Block

5

DATA PRIVACY AND ANALYTICS IN VARIOUS BUSINESS AREAS

UNIT 15

HR Analytics in HR Planning **1-36**

UNIT 16

Data Analytics for Top Management Decision Making **37-61**

UNIT 17

Business and Marketing Intelligence Using Analytics **62-98**

UNIT 18

Data Privacy and Ethics **99-119**

Ref. No. BDCA-SLM-IFHE – 122022 B5

For any clarification regarding this book, the students may please write to The ICFAI Foundation for Higher Education (IFHE), Hyderabad specifying the unit and page number.

While every possible care has been taken in type-setting and printing this book, The ICFAI Foundation for Higher Education (IFHE), Hyderabad welcomes suggestions from students for improvement in future editions.

Our E-mail id: cwfeedback@icfaiuniversity.in

Centre for Distance and Online Education (CDOE)
The ICFAI Foundation for Higher Education
(Deemed-to-be-University Under Section 3 of UGC Act, 1956)
Donthanapally, Shankarapalli Road, Hyderabad- 501203

BLOCK 5: DATA PRIVACY AND ANALYTICS IN VARIOUS BUSINESS AREAS

Analytics has a very crucial role in present day HR function. HR analytics can help answer different questions such as patterns to be revealed in employee turnover, time to hire employees, investment needed to get employees up to a fully productive speed, employees who are most likely to leave within the year, effect of learning and development initiatives and employee performance. Data-backed evidence helps organizations to focus on making the necessary improvements and take future initiatives. Various analysis available for data analytics need to be understood by all learners. If critical business intelligence can be gained from the collected big data, it will be of great benefit to top management decisions and marketing intelligence. It is also necessary for the learner to appreciate privacy and ethics when dealing with such enormous amount of structured, unstructured and semi-structured data.

HR plays a very crucial role in every organization, be it appraisals, training and effectiveness, recruitment and knowledge management etc. Thus analytics in HR is essential to know and implement. Unit 15 – *HR Analytics in HR Planning* - covers introduction to HR analytics, HR analytics and strategy, HR analytics in HR planning and forecasting, HR decision-making and HR analytics for future. It is a vital unit for all learners interested in HR functions.

Top management has to make very crucial decision related to products, finance, infrastructure, customer satisfaction, distributors etc. When real data and related analytics are supplemented to their decision process, it leads to more optimal solutions. Unit 16 – *Data Analytics for Top Management Decision Making* - deals with classification, predictive, cluster and association analysis. It also covers RFM (recency frequency monetary) analysis and market basket analysis (MBA). In addition, this unit covers all the different analysis techniques to be used for analytics like correlation, regression, multiple linear regression, exploratory factor analysis (EFA), principle factor analysis (PFA) and confirmatory factor analysis (CFA).

Marketing is a resourceful function giving many leads to both the management as well customers on products etc. Market research and marketing intelligence, when integrated with business decision making, can lead to more practical and profitable solutions. Unit 17 – *Business and Marketing Intelligence Using Analytics* - spans over data warehousing, need for business intelligence, components, architecture, methodologies and tools for business intelligence. The unit also covers data mining techniques, market intelligence and decision making, Hadoop, introduction to Google Big Query, Apache Spark and Google Dataflow.

Finally, when organizations need to deal with lot of collected real data, enough care needs to be taken about its privacy from outsiders and also needed training is to be given for insiders for ethical usage of the data collected. Unit 18 – *Data Privacy and Ethics* - encompasses the privacy landscape and great data utilization. Other points covered include preferences, personalization and relationships; rights and responsibility; conscientious and conscious responsibilities; data anonymization and balancing the counter intelligence.

Unit 15

HR Analytics in HR Planning

Structure

15.1 Introduction

15.2 Objectives

15.3 Introduction to HR Analytics

15.4 HR Analytics and Strategy

15.5 HR Analytics in HR Planning and Forecasting

15.6 HR Decision-making and HR Analytics

15.7 HR Analytics for Future

15.8 Summary

15.9 Glossary

15.10 Self-Assessment Test

15.11 Suggested Readings/Reference Material

15.12 Answers to Check Your Progress Questions

"HR will not be replaced by data analytics, but HR who do not use data and analytics will be replaced by those who do."

- Nadeem Khan, Managing Director of Optimizhr Ltd.

15.1 Introduction

To get better insights into strategic impact of HR interventions on business, HR analytics are needed. But it has to be remembered that HR analytics only provides the insights, but it is for the HR managers to take decisions.

In the previous unit, we have studied about measuring human resource effectiveness. In order to manage human resources, measuring its effectiveness is important. This helps the organization in knowing the value addition from each department or team. The parameters of HR effectiveness and methods to measure HR effectiveness are also discussed.

In this unit, we will learn about Human Resource (HR) analytics. HR analytics, essentially, refers to application of huge data to an analytical process in the HR department of an organization. The overall objective is to improve and enhance employee performance. HR analytics helps the organization to optimize human

resources and thus contribute to better return on investment (ROI). HR analytics is not only gathering data on employee efficiency, but also providing qualitative insights into the HR processes and using them for the decision-making. HR analytics correlates business data and people data to draw inferences and reflect on possible impact of HR on the bottom-line of the company. HR analytics is all about establishing a cause-and-effect relationship between people and business. Based on analytics information, companies develop business strategies to maximize the gains for both the customer and the business. The aim of HR analytics is to improve employee performance for organizational growth.

15.2 Objectives

After going through this unit, you should be able to:

- Describe HR analytics and explain how data is managed and interpreted
- Discuss HR analytics and strategy and the process of decision-making
- Identify the role of HR analytics in HR planning and forecasting
- Describe the interrelationship between HR decision-making and HR analytics
- Identify the possible uses of HR analytics in the future

15.3 Introduction to HR Analytics

When HR managers demand more involvement in business decisions, they are often asked to produce evidence of enhanced performance or the strategic advantage achieved by investing in human resource. Organizations often term people as its key resource. However, assigning a value to this asset or estimating its appreciation is not an easy feat. HR analytics can be of immense help to organizations in this regard.

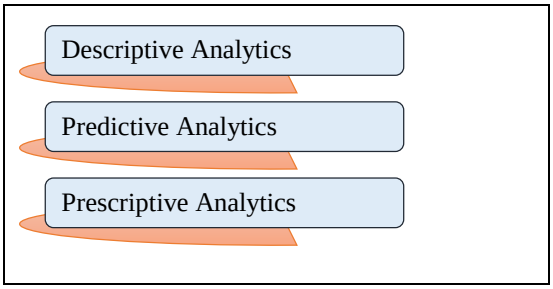
15.3.1 Concept and Definition of Analytics

Across the globe, organizations are shifting focus to evidence-based management practices. Evidence-based management is premised on effective use of data to support evidences, leading to decisions that benefit the organization and its stakeholders.

Analytics is defined as scientific data manipulation. Business analytics is scientific data manipulation for better decision making in businesses. Business analytics makes use of mathematical and statistical techniques and is primarily used in operations research, economics, marketing, financial management and so on. Big data enabled business analytics processes huge amount of data to help in better decision making. In a technology driven era, many decisions like that of operations can be predictive.

Analytics can be broadly classified into three distinct segments – descriptive, predictive, and prescriptive. Let’s take a look at each of these (Refer Figure 15.1).

Figure 15.1: Classification of Analytics



Source: ICFAI Research Center

- **Descriptive Analytics**: Here historical data are taken into consideration for identifying patterns and trends of the behavioral variables.
- **Predictive Analytics** is the next stage of analytics. Here, data is analyzed to predict future behavior. Predictive analytics answers the question of what is likely to happen.
- **Prescriptive Analytics** is the last stage, where the predictions are used to prescribe (or recommend) the next set of actions.

Exhibit 15.1 gives an example of prescriptive analytics.

Exhibit 15.1: Example of Prescriptive Analytics

Google's self-driving car is a perfect example of prescriptive analytics. It analyzes the environment and decides the direction to take, based on data. It decides whether to slow down or speed up, to change lane or not, to take a long cut to avoid traffic or prefer a shorter route, etc. In this way, it functions just like a human driver by using data analysis at scale.

Source: https://www.newgenapps.com/blog/descriptive-vs-predictive-vs-prescriptive-analytics

Predictive decision-making is a more holistic process that is capable of assessing the decisional outcomes, in the beginning itself. It helps managers to calibrate their decisions and to minimize any adverse effect to the outcomes. Over the years, business analytics tools have also become more subject and business function specific. Human Resource (HR) analytics is one such tool.

15.3.2 Meaning of HR Analytics

Traditionally, human resource management was seen as an art, based heavily on gut feelings or intuition to manage effectively. With the advent of technology, the importance and potentiality of data in decision-making was realized. This gave a data-based objectivity in people related decision-making. HR decisions in an organization are varied in terms of scope and complexities. This scientific approach to human resource management in organizations has given birth to HR analytics. However, HR analytics is at an infant stage.

It has been possible to use data meaningfully and able to predict future too.

HR analytics makes use of statistical models to predict the future. It can make predictions based on current HR trends. It aids decisions on critical issues facing the organization like increased rate of attrition, performance appraisals, and so on. HR analytics makes use of statistical and research design for specific HR decisional issues. Relevant data, both within and outside the organization are used for predictions.

Analytics is often confused with metrics. Analytics gives deeper insights for the decision-making process. Analytics enhances the power of the data and encompasses all managerial functions.

Analytics is more like a "continuum", in which one end will have basic ratios and metrics and at the other end has complex algorithm-based predictive analytics. The organization can lie anywhere on the spectrum of the maturity, in terms of the HR processes, quality of data or available capabilities.

With predictive analytics, HR decision-making has become more holistic and it is possible to assess decisional outcomes right at the beginning. It has the capability to manipulate the big data and assess decisional impact before its occurrence. Hence the decision-makers can alter, change or intervene beforehand to improve decisional outcomes.

15.3.3 Importance and Significance of HR Analytics

HR analytics provides scientific facts pertaining to human resource of an organization.

HR analytics starts with the identification of the key business concerns, focuses on strategy and long-term sustainability issues. This is followed by assessing the set of existing knowledge and competencies in the organization. Any gaps are to be met through appropriate interventions. HR managers decide appropriate interventions in advance to improve decisional outcomes.

Many organizations initially begin with descriptive analytics, i.e., making use of metrics to understand the current situation in an organization. Gradually, statistics and research methodology techniques are used for better analysis of decisional outcomes. Finally, HR manager makes use of HR analytics to predict decisions, and thus influence the overall organizational goals.

HR analytics holds lot of importance for organizations. For example, HR analytics is useful in retention of talented employees by using predictive analytics in finding reasons for lower productivity, improving recruitment, etc.

Managers need to understand that the impact of HR activity cannot be assessed in silos, rather it requires cross-functional knowledge. With HR analytics, it becomes possible to analyze the repercussions of an activity.

15.3.4 Benefits of HR Analytics

Earlier, HR managers measured terms like rate of absenteeism, attrition, cost of compensation, etc. These were not sufficient for efficient decision-making needs of the HR manager. With the help of HR analytics, managers can now assess employee engagement, predict the future requirements, and assess the customer relationship management practices.

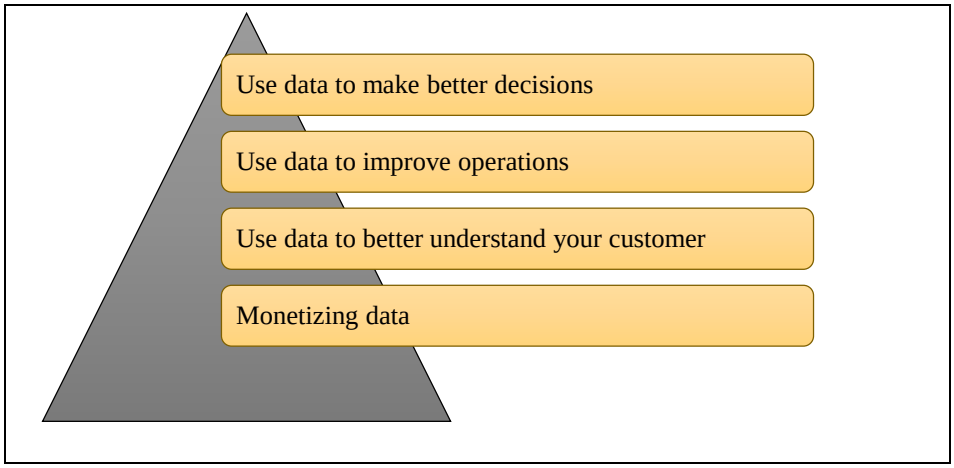
HR analytics helps to predict the trends that help managers take wise decisions by evaluating alternatives to select the best option based on associated big data sets.

It helps managers predict which employees are likely to quit and those likely to remain. It helps in succession planning by determining how organization should strategically do succession planning. HR analysis also helps in scenario planning which is envisioning the future changes and drawing future strategic interventions to correct deviations, if required.

Hence, HR manager can gain an insight into HR functions and can better align them with organizational goals. The HR programs and plans can be better aligned to achieve the strategic intent. Oracle, SAP, and IBM are the top vendors for HR analytics. Apart from this, a number of small vendors also exist. As per the need of the organization, the right vendors can be chosen and the product customized.

Thus, benefits of HR analytics are as follows:[1] (Refer Figure 15.2).

Figure 15.2: Benefits of HR Analytics



Source: ICFAI Research Center

i) Use data to make better decisions - Data-driven HR helps make HR smarter in every possible way. HR analytics can help HR manager make decisions

[1] Marr Bernarad (2018). Data-Driven HR: How to Use Analytics and Metrics to Drive Performance. *Kogan Page Limited*-2018-ISBN 978 0 7494 8246 6. https://books.google.co.in/books?id=rSRTDwAAQBAJ&printsec

about various HR associated functions like strategically aligned recruitment and training and performance appraisal. HR data can be used for decision making, within the department or used by top management for some other decision-making with HR team's support.

ii) Use data to improve operations - HR analytics can help improve efficiency in HR operations. The data gathered about the way HR is functioning can unlock certain grey areas that need to be worked upon. It provides clues about HR functions like safety and well-being of employees or recruitment. It helps to answer questions like "Where do we spend most of our time on?" and "How do we streamline and improve these functions?" Analytics helps to automate or standardize certain processes and determine ways to improve them.

iii) Use data to understand your customer better - This is one of the most publicized areas of analytics. Analytics helps better understanding of the customers' needs and behaviors, their preferences and the level of satisfaction. The employees of an organization are its the internal customers. Hence, HR analytics help better understanding of the needs and demands of the internal customers of the organization.

iv) Monetizing data - Monetizing data creates new revenue streams for a company. However, organizations may decide to monetize data from other areas of business impacting the HR. HR analytics, in such scenario, can be used to identify changes like new skillsets that company needs to recruit or the value that HR can add to the overall organization's changing strategy.

Example: Tata Steel Uses HR Analytics to Democratize Access To People Data so that Better Decisions can be taken at all Levels

Tata Steel has realized that insights from HR data analytics are made available to everyone in the HR department so that these insights are embedded into decision making at every level. For this, the company analytics team is closely working with the HR managers and also enhancing the HR analytics skills of the HR staff. The idea is to eliminate bias in decision making which is due to not having HR analytics insights.

Source: The Business Value of Democratising Access to People Data at Tata Steel | myHRfuture Date 14/02/2022, Accessed on 27/09/22

15.4 HR Analytics and Strategy

HRM has evolved from being a traditional administrative job to a strategic function. HR uses technology and analytics for its functioning. HR analytics helps to reinforce positive outcomes of talent management, performance improvement, employee engagement and so on. With the use of analytics, HR is able to measure its value and, in the process, facilitates understanding of how human resources

align with organizational strategies. For example, with the help of HR analytics, it is possible to track not only cost of recruitment but also important details, which may have significant impact on business and organizational strategies like cycle time to hire, cost of training and on boarding, diversity inclusion, culture fit and so on.

Cause and effect relationship between HR functions and business goals help reduce the bad investment in HR initiatives. For example, training programs with poor transferability, wheretraining does not bring any incremental change, can be stopped.

Check Your Progress - 1

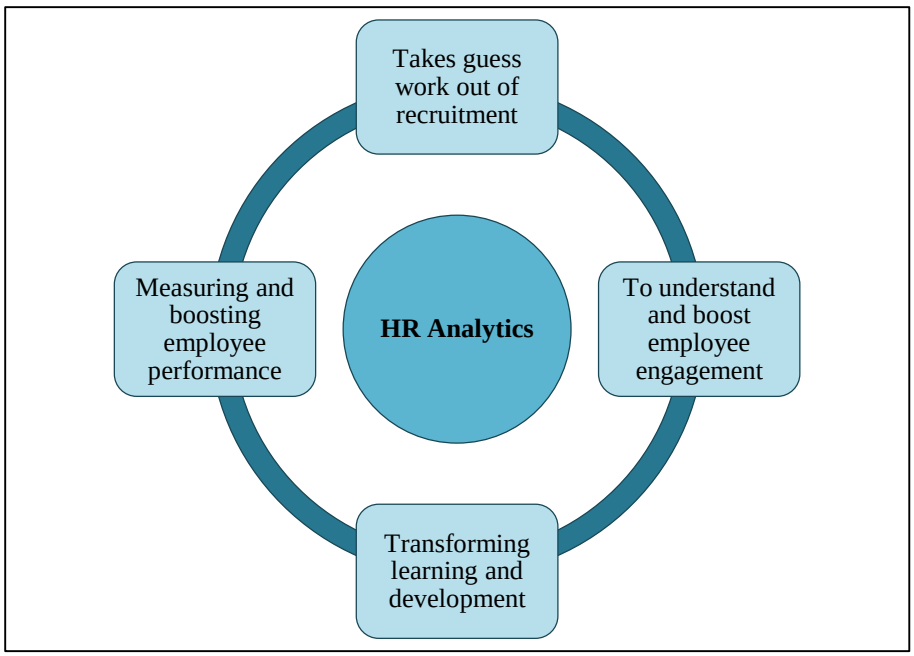
1. What is the main purpose of business analytics?
    a. Higher profit
    b. Operational efficiency
    c. Restructuring business
    d. Better decision-making
    e. Investment purpose
2. Which of the following statements is correct about HR analytics?
    a. It is all about metrics
    b. Predictive analysis is prevalent in all HR analytics
    c. Ratio dominates HR analytics
    d. Organizations lie anywhere on the HR analytics spectrum, depending on the maturity of the HR processes, quality of data, and available capabilities
    e. Organizations lie only in some specific position on the spectrum on the maturity of the HR processes, quality of data, and available capabilities

15.4.1 Aligning Human Resources to Business through HR Analytics

Using predictive modeling, HR analytics can align business and organizational strategies with manpower planning, talent management, change management, redundancy planning and so on. Manpower or HRP ensures the availability of right manpower at right time, based on current trends and future business goals.

The analysis of the demographic data of the current employees helps to prepare for the future manpower needs of the organization. Alignment of human resources to business through HR analytics can be achieved, in the following ways (Refer Figure 15.3).

Figure 15.3: HR Analytics



Source: ICFAI Research Center

Takes guess work out of recruitment: Recruitment is an expensive affair for an organization and impacts its performance in the long run. Guess work or gut feeling does not work often and can cost companies greatly. Category-wise manpower requirement can be assessed and accordingly plans for new recruitment can be made. A data driven approach to recruitment helps companies find employees, who are well suited to the needs of the organization. Any HR dashboard template provides practical insights about existing professionals, their occupancy, performance levels, which help the HR and management in general to plan for the required manpower based on annual strategic plans of growth. These are crucial during M&A, laying off people, competitive assignments, planned growth, change management, etc. Take for example, an IT company is bidding for a new agile project in Python with reference to recruitment. Some of the people offered letters of selection have Python experience. A good dashboard, if properly managed, can help organizations to have efficient workforce as also plan for other necessary resources with required experience.

To understand and boost employee engagement, analytics driven tools are helping organizations to better understand and enhance the employee experience. Employee engagement and culture is boosted by adopting an integrated focus on bringing together all the HR and management practices driven by the tools of analytics.

As the organization grows, especially with multi facility located operations, analytics helps management to assess HR parameters like employees' work

experience and their current engagements. Availability of manpower at a particular time for various projects can also be ascertained. This can reduce negative sentiments among employees.

Transforming learning and development: Online learning is becoming a prominent part of how organizations develop their resources.HR analytics helps understand the learning pattern of employees and cater to their unique way of learning by suggesting right contents. It helps identify the knowledge and skills required in the future and train people to make them future ready.

With the availability of a number of online learning platforms like MOOCs and SWAYAM, professionals are getting certified on their own. Organizations need to take note of this aspect while deciding on the training programs. Analytics in this direction helps organizations to track the knowledge level within the organization. This is matched against the needed skills to take decision on the training vendorsand the training calendars.

Data analysis helps to measure employee performance more accurately and review performance in a smarter and agile way. Several parameters can be studied to help better understand what motivates an employee and what does not. Implementing an incentive scheme and its impact on employee morale can be studied and monitored. It also helps to understand employee's performance and satisfaction level and the cascading effect it has on customer's satisfaction.

Any time, a data based feedback to employee gives more value than a mere feel statement. Automated HR metrics help compile individual performance data at shorter intervals and help to analyze these data for useful metrics in helping and boosting employee performance on periodic basis at short intervals. These can also be made online.

Example: Bank of America Deploys data HR Analytics to Ensure Employee Growth in Customer Relationship Management

Bank of America noticed low employee morale and productivity. HR analytics was done on HR data and the data provided insight that the low morale and productivity was a result of higher stress levels and negative experiences at the workplace. The company implemented a strategic HR initiative centering around "collective breaks". During collective breaks, the team members can discuss and exchange tips for handling stress generated during customer interactions. The new initiative resulted in 23% increase in performance and 19% decrease in stress levels. This resulted in the growth of employees in the customer relationship management which is aligned with the company strategic objectives and goals.

Source: 3 Examples of How Effective Usage of HR Data Caused Employee Growth - Agile HR Analytics (agile-hr-analytics.com) date 02/04/2022, Accessed on 27/09/2022

Activity 15.1

Consider you are the human resource manager in your organization. Prepare a four point agenda to convince the top management to adopt analytics in the field of human resource management?

Answer:

Check Your Progress - 2

3. What would you eliminate to align human resources to business through HR analytics?
    a. Understand and boost employee engagement
    b. Transform learning and development
    c. Retain intuition or guess work in recruitment
    d. Measure and boost employee performance
    e. Understand learning pattern of employees
4. Which of the following statements cannot be associated with HR analytics in the present time?
    a. Assess employee engagement
    b. Predict the future requirements
    c. Assess the customer relationship management practices
    d. Gain insight into HR function
    e. Just measure some of the terms like rate of absenteeism
5. Which of the following statements uniquely correlate to predictive analytics?
    a. Here, historical data is taken into consideration for identifying patterns and trends of the behavioral variables
    b. Here data is analyzed to predict future behavior and explains what is likely to happen in future
    c. Here, the predictions are used to prescribe (or recommend) the next set of actions
    d. Use of mathematical and statistical techniques
    e. Measure some of the terms of HR

15.5 HR Analytics in HR Planning and Forecasting

An unprecedented increase in the number of job openings indicate that companies need to better plan and recruit the human resource. HR analytics has emerged as a key focus area to better plan and forecast human resource to match the organizational requirements.

15.5.1 HR Forecasting

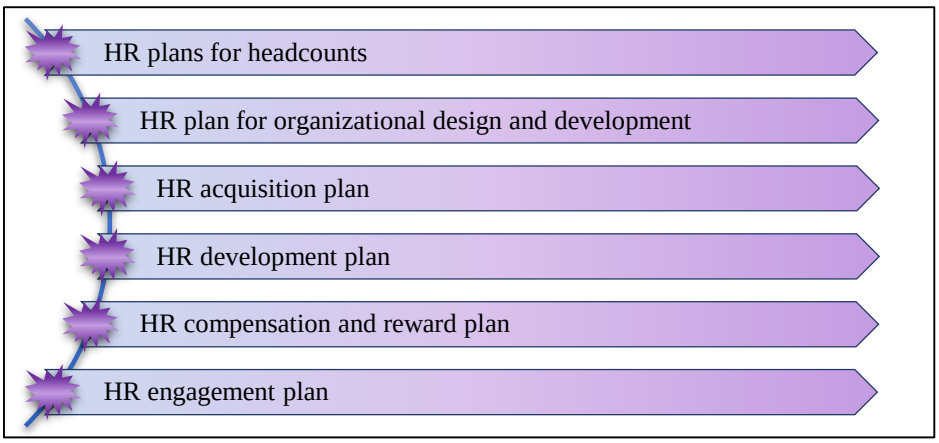
HR forecasting focuses on measuring the implications of human resources on organizational strategy. HR forecasting is done after thorough analysis of the economic, technological and organizational forces on the human resources. This calls for some structured steps such as framing a business strategy, relating business strategy with various HR scenarios, assessing demand and supply of human resource and translating HR forecast into HR plan.

15.5.2 Components of HR Plans

HR plan is the end result of HR forecasting and once the plan is ready, it has to be executed. With HR plan, HR managers weigh various strategic options, analyzing various HR data and information on culture, training and learning. Also, HR plan needs to be dynamic as it is based on organizational strategy and is influenced by environmental changes. HR analytics brings a feature of predictability.

HR plan is inseparable from organizational business plan and has to be systematically reviewed and changed. The components of HR plans (Figure 15.4) include:

Figure 15.4: Components of HR Plans



Source: ICFAI Research Center

i) HR plans for headcounts: It is a quantitative plan in which the headcounts are optimized as per the business strategy.

ii) HR plan for organizational design and development: Such a plan considers organizational structures for both the current and future ones.

iii) HR acquisition plan: This plan helps to frame recruitment strategies which extend to any acquisition, redeployment, retention plan, and so on.

iv) HR development plan: It helps plan for the training and development programs to assess the return on investment from employees' training activities, and so on.

v) HR compensation and reward plan: This helps to compensate employees and reinforce any positive performance. It is instrumental in keeping the employees happy and satisfied.

vi) HR engagement plan: It helps to ensure the engagement of human resource with organization and involve attitudinal survey.

There may be various other types of HR plans which an organization may develop as per its nature and strategic requirements. With effective HR plans and forecasts, organization can optimize its performance and ensure high degree of commitment from human resource.

15.5.3 KPIs for Employees

Current practices in HR analytics include defining KPI (Key Performance Indicators) for employees and relate measurements to ensure that these are met.

Some examples of KPI for HR can be[2]:

- **Employee productivity rate:** Organizations need to define these based on their work culture, nature of projects and the nature of products and services. It helps in assessing the capacity of growth in terms of production of human capital.
- **Employee satisfaction index:** While time of appraisals is one source, employee satisfaction can also be measured through employee attitude and engagement surveys. It is obvious that dissatisfaction is the main cause for employee turnover.
- **Employee engagement index:** This is linked to productivity. Employee engagement is measured through projects online, completed projects in recent times and foreign assignments, besides others. High employee engagement leads to higher productivity, lower turnover, better customer service, and many other positive outcomes.
- **Employee innovation index:** Innovation is also measured through new products launched, processes modified, waste elimination and response to kaizens etc. Innovation is the key driver of business success and HR has a major role to enable this innovation.

[2] https://www.analyticsinhr.com/blog/human-resources-key-performance-indicators-hr-kpis/

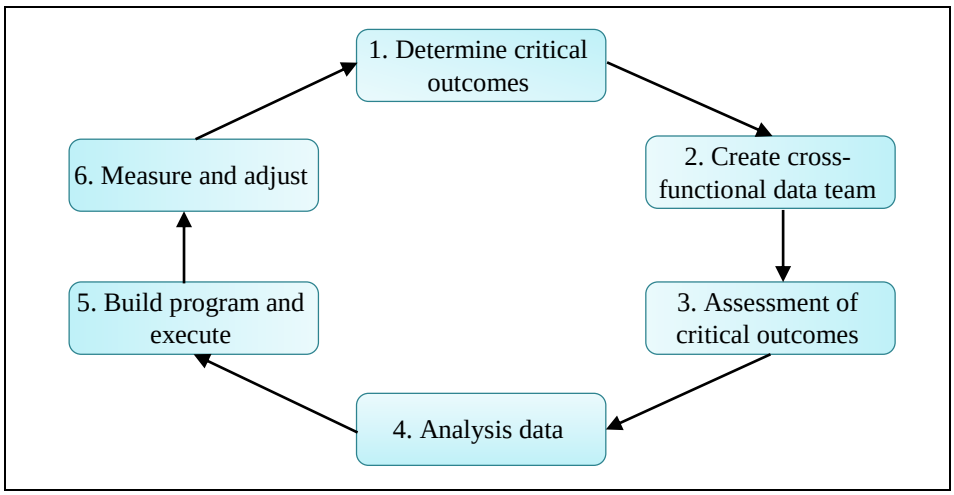
- **Internal promotion rate:** This KPI is a division of a number of major functions that were filled through internal promotion divided by the total number of senior positions filled. Internal promotions are better, as they understand culture better, speed up faster, reduce the risk of a likely bad hire and assure longevity.
- **Net Promoter Score:** A Net Promoter Score (NPS) is the right way of measuring to what degree people would recommend a service or business to another person. Depending on the strategic goals, NPS is an important KPI for HR.
- **Quality of hire:** The quality of hire is the percentage of new hires that are given a good rating by their manager during their periodic performance review, especially the first one. Quality of hire indicates HR recruitment process. Maintaining a high quality of hire rating helps the organization to reach all its strategic goals.
- **Turnover rate:** Turnover is a necessary common metric and an important KPI. High turnover is costly in many aspects, and slows down the progress, reducing the productivity.
- **Training effectiveness:** Many training programs may be organized for the benefit of individual employee leading to the productivity and growth of the organization. Thus, metrics need to be in place for measurement of training effectiveness.

15.5.4 Steps for Conducting HR Analytics for Strategic HRP Model

A six-step process for conducting HR analytics for strategic HRP model was suggested by Scott Mondore, Shane Douthitt and Marisa Carson.

Refer Figure 15.5 for six-step process of implementing HR analytics.

Figure 15.5: Six steps for Implementing HR Analytics



Adapted from Banerjee, P., &et. al. (2019). Practical applications of HR Analytics. Sage Publications India Pvt. Limited.

The steps of implementing HR analytics to support the HRP model, include:

i) Determination of the critical outcomes - An organization needs to identify the critical outcomes that it need to focus on. These outcomes are highly influenced by an organization's vision and mission statement. The productivity needs to be closely linked to the needs of the organization.

ii) Creation of Cross-functional data team - It aims to identify the key personnel who own the data. There are certain people who hold a lot of data and are one point contact for access to that data. It's a good idea for HR analysts to get the data required for their task. The head of the department or the line managers would be key contact persons many a time. Lot of inter-linking of data is required to infer fruitfully from the data.

iii) Assessment of the critical outcomes - It is used to understand the procedure or methodology to capture the current employee data. It involves the periodicity of measurement, the level of measurement and so on. It is important as there has to be a common base in order to have a meaningful comparison of data over time.

iv) Analysis of data - The knowledge of the advanced statistical techniques and tools for analysis of data saves a lot of time in calculating the results which otherwise would be a tedious task. Relating HR initiatives with business outcomes, prioritizing the type of interventions, and calculating the expected return on investment are of paramount importance for HR analysts.

v) Building and executing a program - Here, the HR analyst builds an action plan based on the results gathered from the analysis of the data. If required, the prepared plan must be executed. Analysis of data helps understand the factors that are affecting the performance. The magnitude and direction of such relationship would allow HR manager to make informed decisions. The outcome may be positive and significant or positive and insignificant or negative and significant or negative and insignificant. The negative and significant factors act as detrimental to the organizational performance whereas the positive and significant factors are most important to contribute towards organizational outcomes.

vi) Measurement and adjustment - The sensitivity and actual outcome of the analytics process is taken care of at this stage. The measurement scale is improved, if found necessary and modified to suit the ongoing market requirements. This makes HR analytics a cyclic process as feedback is incorporated and scales reworked as per the inputs. The process is sequential and if the intermediate steps are overlooked, the results may not be correct. The activity, being periodic in nature, calls for an efficient system in place, to do the needful and store information in a way that may be required by the ever advancing statistical tools.

Other areas of HR analytics are:

Turnover:

- Analyze collected past data on turnover to identify trends and patterns and identify the reasons why employees quit.
- Collected data on employee behavior related to productivity and engagement. This is analysed to better understand the status of current employees.
- Correlate both types of data to understand the factors that lead to turnover.
- Help create a predictive model to better track and flag employees who may fall into the identified pattern associated with employees that have quit.
- Develop strategies and make decisions that will improve the work environment and engagement levels.
- Identify patterns of employee engagement, employee satisfaction and performance.

Recruitment:

- One has to ensure fast, automated collection of candidate data from planned and available multiple sources.
- Consider candidates' extensive variables, like developmental opportunities and cultural fit.
- Identify candidates having attributes that are matching the top-performing employees in the organization.
- Avoid habitual bias and ensure equal opportunity for all candidates; with a data-driven approach to recruiting, the viewpoint and opinion of one person can no longer impact the consideration of applicants.
- Compile metrics for how long it takes to hire for specific roles, once the need is identified within the organization, helping departments to be more prepared and planned when such need to hire arises.
- Present historical data related to periods of over-hiring and under-hiring. This enables the organizations to develop sustained long-term hiring plans.

Refer Exhibit 15.2 for a case study on Lowe' usage of HR analytics data.

Exhibit 15.2: Case Study: Connecting HR Data and Business Outcomes at Lowe's

Lowe's is an American company which operates a chain of retail home improvement and appliance stores. Since 2007 the company has been using a data-driven HR business model to highlight the connections between HR decisions and business outcomes. As always, the first hurdle was to build the necessary business logic to establish management belief in the project across various departments.

Contd….

Lowe's analytics team sought to link various HR data (such as engagement, turnover and sick time) to marketing data (such as customer satisfaction and loyalty), operations data (such as shrinkage, which measures inventory loss resulting from factors such as employee theft and fraud) and financial metrics (such as sales per square foot and net income before tax). Once leaders arrived at a consensus on the business logic behind HR policies, a cross-functional team comprising finance, marketing and operations, facilitated and supported by the human resources division, was established. The result was a set of statistical models that were built with inputs from across the company and were therefore accepted by all major stakeholders.

Results and impact: One of the major findings from the analysis was the relationship between employee engagement, customer satisfaction and the impact on revenue. Lowe's intuitively sensed that there was a connection between engagement and customer satisfaction. Through the analytics exercise, the company was able to attribute a monetary value to this linkage. One example of a quantitative conclusion was the relationship between engagement and average ticket—the amount of money a customer spends per transaction. By encouraging greater dialogue (including asking customers a specific set of questions) between employees and customers, customers felt they were having a better store experience and spent 4% more. Lowe's reached a conservative estimate that the gap between the highest and lowest engaged stores constituted more than US$1m in sales per year. "After measures and analytics, comes the process" says Wayne Cascio, professor of management at the University of Colorado-Denver.— Process is using the results to make real decisions which impact the business. Management shared the results with the grass roots of the company, improving the chances that systemic change would be deeply rooted. As a result, focus on employee engagement has spread throughout the entire organization. Management teams appreciate the value of employee engagement and are keen to learn how engaged their own teams are and what they can do to boost engagement. In this way, workforce analytics has achieved genuine, sustainable change with measurable business outcomes that have resulted in competitive advantage for the firm.

Source: http://c.ymcdn.com/sites/www.hrps.org/resource/collection/86D817D1-E244-4847-A103-BC7E19E57AB6/HRPS_PS34-2_Final.pdf

Example: LA County (USA) deployed HR Analytics to Improve Recruitment Efficiency by Reducing The Average Number of Days to Hire a New Recruit

Los Angeles County had an inefficient recruiting process that took an average of 384 days to hire a new recruit. The HR department used HR analytics to improve recruitment efficiency. Before the new system the analytics team had faced challenges at every stage of data analytics process, data collection, analysis and sharing the insights.

Contd….

The right tools were not there. The data was distributed across the HR department servers and sharing was also a challenge.

The two main objectives were a. Find the bottlenecks causing the delay, and b. Enhance the experience of the potential candidates. If the experience is not good and the process is taking long time , better talent will move on to other employers. With the interactive dashboards available as a result of the new analytics platform, the HR team could identify the blocks, eliminate them, and speed up the process

Source: LA County modernizes recruiting with Azure and Databricks (edwardsbrandtiowarealty.com) date July 06, 2022, Accessed on 27/09/2022

15.6 HR Decision-making and HR Analytics

HR analytics is believed to significantly alter and improve the HR decision-making, bringing a holistic change in the organization. Some HR managers still find it an over-pitched promise. HR analytics can establish a link between HR decisions and employees and can improve decision-making. HR analytics helps diagnose an HR issue. It can facilitate taking preventive actions by making appropriate interventions.

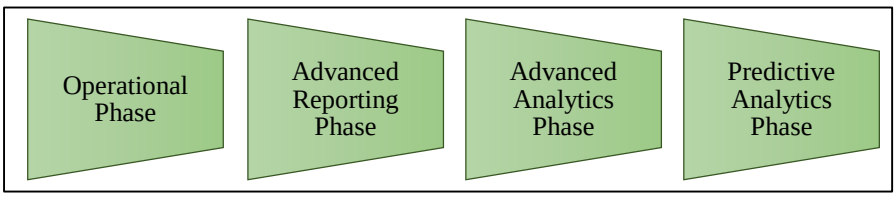
We will now discuss the Phases of Development of HR Analytics for HR decision-making, descriptive HR decision-making, correlational HR decision-making, predictive HR decision-making and critical HR decision-making.

15.6.1 Phases of Development of HR Analytics

Different phases of HR analytics are described using the Bersin by Deloitte's maturity model of talent analytics[3]. The model lists out a scientific progression from operational reporting to predictive analytics by using four different phases or levels.

The four distinct phases are as follows (Refer Figure 15.6).

Figure 15.6: Different Phases of HR Analytics



Source: ICFAI Research Center

i) Operational Phase: This is also called reactive phase. This involves the task of operational reporting on performance issues like compliance matters and so on.

[3] https://www2.deloitte.com/au/en/pages/human-capital/articles/getting-started-talent-analytics.html

ii) Advanced Reporting Phase: This phase is also known as proactive operational reporting phase. Here, a multi-dimensional analysis is performed by using dashboards and others alike. At this stage, organization can also go for benchmarking.

iii) Advanced Analytics Phase: This phase is also known as proactive identification of problems or decisional issues for reaching an actionable solution. In this phase, statistical modeling techniques, root cause analysis and the like are performed for solving business issues. From this phase, we actually begin to use HR analytics solution to solve our business problems.

iv) Predictive Analytics Phase: In this phase, predictive models are developed, risk analysis is done, scenario plan developed and so on. At this phase, ultimately HR functions focus on how to derive the benefits of HR analytics.

Exhibit 15.3 below is an example of using HR analytics to reduce employee turnover.

Exhibit 15.3: How Credit Suisse Used Predictive HR Analytics to Reduce Employee Turnover

The investment banking major, Credit Suisse, deployed predictive analytics to identify employee churn and determine the reasons behind employees wanting to quit. This information was anonymously shared with line managers to help them reduce turnover risk factors and retain their talent better.

Based on these insights, Credit Suisse also provided special managers with training on retaining high-performing employees who were likely to give notice. The bank saved an estimated $70,000,000 a year in recruiting and on boarding costs as a result of this initiative.

Source: White Danny (2019). Top 3 Examples of Predictive Analytics in HR. TechFunnel.com- June 5, 2019.
https://www.techfunnel.com/hr-tech/top-3-examples-of-predictive-analytics-in-hr/

Hence, knowledge of strong statistical tools and skill of data analysis are highly desirable for an HR analyst.

Some statistical tools are discussed below:

Regression analysis: As one of the most common tools of statistical analyses, regression is used to capture the relationship between one or more context variables and an outcome in a function. The goal here is to predict the future progression of the outcome based on values of the context variables.

Classification analysis: One useful example of classification analysis would be to predict the success rate of a team based on formation of the team composition and other context variables. Organizations build project teams based on experience, availability, needed technology, domain knowledge and past

individual performance. It is more valuable to give emphasis to other factors like role preference, team size, background, leadership style, team dynamics, assignment duration and budget. This calls for a reliable, in-house dataset in order to train the model. The classification techniques aid in predicting the right team composition which will have the highest success rate.

Clustering analysis: Clustering is a technique to describe data and to find general patterns. It is used when available data are not, or ambiguously, labeled. Clustering works by finding observations that are similar to each other. These observations are then 'clustered' so that the groups can be labeled and categorized.

Association analysis: This technique could be used to identify patterns in HR practices such as on boarding, career paths, education and talent management. It is also used to identify which patterns are associated with happy and productive employees. It could then be used to give feedback to an HR system for customized content. Much like the way Amazon and Netflix offer customized content to consumers.

Anomaly detection analysis: Accidents at workplace are often the result of fatigue and long working hours. A review of 12 studies showed that employees working over 12 hours per day had a 38% higher risk of occupational injury than those working 8 hours a day.

Working 10 hours per day increased the risk of injury by 15% compared to working 8 hours per day. Thus anomaly detection analysis is aimed at helping in identifying those employees who work longer than a specified threshold, more so in high-risk occupations. These are construction, manufacturing and engineering. This helps in prevention of accidents and injuries at the workplaces.

15.6.2 HR Decision-Making

Management practices have undergone tremendous changes. Using data to ascertain the evidence for decision-making has become the norm in most of the organizations. The concept of evidence-based management practices was coined by Pfeffer and Sutton in 2006 and Briner in 2009. Evidence based decisions add value to organizations. They are futuristic and sustainable.

If HR decision-making is not carefully done, several problems would arise. For instance, the selected candidates may become unproductive, unable to cope in a team or they may fail to integrate themselves with the organizational culture. As such, HR decision-making is vital because it is based on the judgment of HR managers. HR decision-making, as such, is defined as HR manager's judgmental thoughts on action[1].

Analytics-based HR decisions reduce decisional bias because they are supported by data. HR decisions can be classified into two types- financial HR decisions and data driven HR decisions. Financial HR decisions are about ROI on any

HR functions such as training and development. Data driven HR decisions are those, which are facilitated by analytics.

However, there are few constraints also for making effective decisions. Organizational systems, regulatory influences, social impact and responses of stakeholders influence decision-making of HR managers.

Figure 15.7 illustrates different types of HR decision-making.

Figure 15.7: Types of HR Decision-Making

Source: ICFAI Research Center

15.6.3 Descriptive HR Decision-Making

Descriptive HR decision-making process uses HR metrics and Human Resource Information Systems (HRIS) to get insights on decisional issues, on the basis of which decisions are taken. For instance, high attrition may be due to lower compensation or lower perks and facilities offered. As per the descriptive decision-making process, in case of high attrition, organizations resorting to conventional methods may check the past records and instances and take measures to reduce attrition. Descriptive HR decision-making process is normally reinforced by correlational HR decision-making.

15.6.4 Correlational HR Decision-Making

Correlational HR decision-making, an analytical technique is very useful for decision-making. This technique is used to know the relationship between two or more variables. *HR* Analytics explains the *correlation* of people data with business data and its impact on the performance of an organization.

Correlation explains the following points:

- When there is no relationship between the variables, it means that a change in one variable will not impact the other variable.
- When relationship exists between two variables, correlation explains the direction of the relationship. The direction tells whether the relationship is positive or negative.
- When it is clear that there exists a relationship between the two variables, the HR manager should know whether the relationship is strong, moderate, or weak.

These three points are vital for decision-making for HR manager. It is important for the HR manager to understand the direction and degree of impact among these variables. This enables the HR managers to take decisions. If the manager finds negative and weak relationship, he/she can alter the methods adopted.

Some parameters studied through correlation include:

- Academic credentials vs output.
- Institute studied vs job satisfaction.
- Manager assigned vs team performance.
- Previous work culture vs present environment.
- Organization goals clarity vs results.
- Motivational parameters existing vs results /attrition.
- Change management vs results.

However, correlation does not explain cause and effect of relationship; it only explains the association between the variables. When we look at the why of it, that is the causal relationship, HR analytics needs to be used (Refer Exhibit 15.4).

Exhibit 15.4: HR Insights through HR Reporting and Analytics

HR reporting and analytics team can partner with the HR function to provide insights, which have helped to develop more impactful HR processes, and help deliver greater outcomes for the business. Many organizations use the engagement data for majority of the HR insight to be created. CCE (Coco-Cola Enterprise) used this method to increase the level of insights developed through the method. By using longitudinal data, they tracked the sentiment in the organization. It provides leaders with a good indicator for checking the power of HR initiatives. The question is whether there is correlation between engagement and business results. For CCE, this point turned important to explain the implications of HR data insights to the rest of the business. "There have been a number of examples where we share insights that are being acted upon. One example is, the engagement survey that is run every couple of years.

The survey includes three questions related to communication. The business was keen to analyze the correlation between how an employee scores a manager, in terms of communication, and key performance indicators across the sites. Across all our sites, there was a positive correlation between the communication patterns and strength of the leaders' communication and related business outcomes.

Source: Coca - Cola Enterprises (CCE) Case Study: The Thirst for HR Analytics Grows. Valuing Your Talent

https://www.cipd.co.uk/Images/case-study-coca-cola_tcm18-19987.pdf

15.6.5 Predictive HR Decision-Making

Predictive HR decision-making is based on big data analysis. Two tools that are used in predictive analysis are causation and regression analysis. Causation analysis differs from correlational analysis. While correlational analysis gives the relationships between variables, causation analysis explains the variables that impact decision-making. Regression analysis is a statistical method that enables understanding the intensity and direction of the relationship between two or more variables. Regression analysis determines relationship such as between the sales and profits, over the past several years. The regression results show, if this relationship is significant or not.

Thus, regression analysis helps in arriving at decision-making with precision and accuracy.

Some parameters to study are:

- Employees from same school in new recruitment (based on last year lot from same school) and expected behavior/output.
- Rural vs Urban schools.
- Connection of hobbies for creativity/innovation.
- New training conducted vs probable results.
- Policies vs millennial and other employees.
- Work ambience – suitability to millennials.
- Training conducted – on job performance – improvement-rating training vendors – L and D aspects.

HR Analytics in a way is essentially the correlation of people data with business data to create people strategies based on available information so as to positively impact company performance.

Most ERPs come with their business intelligence or business analytics add-ons. Alternatively, you could opt for standalone tools like Tableau, Visier, Aquire, etc. Once you have data and the tool, you need to start forming your own hypothesis as to what is the cause and what is the impact. Gather all relevant data points and find out the correlation factor to see if your assumption was actually true.

In one of the organizations it was figured out that, if a manager is also on a team member's personal friends list, the employee is more likely to stay back longer. This triggered training programs for managers to develop personal bonds which in turn reduced attrition (measured by month-on-month attrition rates before and after training). One of our Talent Acquisition managers was able to plot the performance of new hires (measured by sales registered in first year) vis-ã-vis their source of hire to conclude that employees hired from Tier 2 B-Schools outperformed those from Premier institutes. Some of them even outperformed their tenured seniors. This predictive analysis helped fine tune hiring strategy, reduce costs and also increase profits.

The idea was to build and present an HR dashboard showcasing the impact of HR on business, establishing a cause-and-effect relationship between what HR does and actual business outcomes.

Companies like Google and Facebook extensively use big data to manage their core business and HR functions.

Exhibit 15.5 shows how IBM is able to predict attrition.

Exhibit 15.5: Employee Attrition is Predictable

IBM used Watson algorithms to predict attrition in their firm by using predictive attrition model. With this model, they could identify who is going to leave, when and why. The model works by clustering/ classifying employee profiles based on various attributes such as age, sex, marital status, education level, work experience, distance from hometown, etc. Watson, thereafter, suggested the steps to be taken to improve engagement and prevent attrition. These were often related to skills development, education, promotion, and raises. These efforts to improve retention saved about US$ 300 million. IBM said the predictions are now in the 95 per cent accuracy "range".

Predictive Attrition Model helps in not only taking preventive measures but also in making better hiring decisions.

Source: Rosenbaum Eric (2019). IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs.
CNBC-April 3, 2019.
https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html

15.6.6 Critical HR Decision-Making and HR Analytics

HR analytics has a major impact on the two HR functions, recruitment and selection. This is mainly because choosing the right candidate and retaining him/her are critical aspects. Analytics help in predicting the retention probabilities of the new hires.

Exhibit 15.6 illustrates the benefits derived by HR teams from workforce analytics.

Exhibit 15.6: 13 Ways HR Teams Can Benefit from Workforce Analytics

Big data has been providing ground-breaking benefits, if organizations knowhow they use that data to highlight insights. Human resources departments are able to use big data alongside workforce analytics for hiring and many other HR functions. Forbes Human Resources Council explores the impact of workforce analytics and big data on the HR department's ability to hire the right people on 13 components.

Contd….

1. **Evidence-based Recruiting (ER) Decisions**

   Modern HR organizations benefit greatly from predictive analytics by eliminating potential misfits at an early stage. They can improve the efficiency of the recruitment process, reduce time, cost, and effort.

2. **Workforce Intelligence for Proactive Strategy**

   Workforce analytics is able to be proactive about the future talent needs of the organization, so as to determine continuous strategic and operational shifts needed for the best talent outcome.

3. **Growth and Forecasting**

   Without the ability to know where you came from, how will you know where you want to go and how to get there? Using big data and analytics will provide a resource and a tool to help formulate plan and path and future direction leading to effective decisions.

4. **Driving Business Strategy**

   As HR and ER departments become more integrated and drive employee experience, sharing insight and trends with decision-makers is vital.

5. **Better Performance Evaluation**

   Analytics and big data have shown great potential in areas such as performance evaluations. Millennials and Gen Z make up the majority of today's workforce, and they prefer real-time feedback by leveraging cloud-based solutions.

6. **Lowering Turnover Costs**

   Predictive analytics can now enable us to build algorithms to predict turnover on an individual employee basis.

7. **Creating Fair Pay Systems**

   One big benefit is creating fairer pay systems by knowing how large companies are paying and comparing with one's own company.

8. **Associated Engagement**

   Similarly, one can understand what benefits will help make better decisions about the future investment. Big data can also help reduce compliance risk.

9. **Talent Discovery and Mobility**

   Understanding data about employees, including skills, experience, performance indicators as well as analysis of trends over time help identify high potential candidates that are a better fit.

10. **Trend Spotting**

    With the help of HRIS, trends in the workplace such as attrition, retention can be analyzed to make employees more engaged.

Contd….

11. Streamlined Talent Acquisition

By leveraging big data, employers can determine what qualities make an employee successful in a role, then filter through thousands of resumes to identify the most qualified candidates.

12. Understanding How to Reskill Candidates

Workforce analytics will play a key role in breaking down job requirements to better understand where and how to reskill candidates or employees.

13. Separating Fact from Fiction

Data analytics can help HR teams distinguish hearsay from reality. Skilled managers can address HR issues with accurate information that can realign teams with their mission and goals.

Source: Expert Panel, Forbes Human Resources Council (2020). 13 Ways HR Teams Can Benefit From Workforce Analytics.

Forbes Human Resources Council-Jan 15, 2020.
https://www.forbes.com/sites/forbeshumanresourcescouncil/2020/01/15/13

Example: Keolis Downer Hunter (Australian Multi Modal Transport Organization) used HR Analytics for Better Decision making related to HR Planning

Keolis Downer Hunter is an Australian Multi modal transport organization operating in Newcastle and Lake Macquarie. The company integrated buses, ferries, light rail and on-demand services. It employs around 4000 employees, and the company was facing the HR challenge of not having insights based on data. The decision making was based on managerial expertise.

The organization went for HR analytics solution with a view to get HR insights for taking decisions related to HR planning throughout the company. The company engaged a HR analytics company Agile HR Analytics (a Microsoft partner) who quickly provided dashboards and reports available to HR at various levels to take more meaningful decisions

Source : Dashboards Case Study | Power BI - Agile HR Analytics (agile-hr-analytics.com), 2021, Accessed on 28/09/2022

Check Your Progress - 3

6. At which step of strategic HRP Model, is sensitivity and actual outcome of the analytics process taken care of?
   a. Determination of the critical outcomes
   b. Creation of cross-functional data team
   c. Analysis of data
   d. Building and executing a program
   e. Measurement and adjustment

7. Which phase of HR decision-making and HR analytics is also known as reactive phase?
   a. Operational phase
   b. Advanced reporting phase
   c. Advanced analytics phase
   d. Predictive analytics phase
   e. Measurement phase
8. Which phase of HR decision-making and HR analytics is also known as proactive identification of problems?
   a. Operational phase
   b. Advanced reporting phase
   c. Advanced analytics phase
   d. Predictive analytics phase
   e. Measurement phase
9. What does Descriptive HR decision-making do?
   a. Uses HR metrics and Human Resource Information Systems (HRIS) to get insights on decisional issues
   b. Know the relationship between two or more variables
   c. Know the correlation which explains the direction of the relationship
   d. Develop scenario plan
   e. Operational reporting

15.7 HR Analytics for Future

It is evident that HR functions in future will exploit Analytics for improved delivery of its services. Future HR managers will be using data and analytics for HR decision-making. For this, in-depth knowledge in HR analytics would be required. Technology enabled HR decision-making will be more effective in driving business outcomes of organizations. HR departments in future would integrate HR analytics with strategic-level business analytics, making HRM also a business function. Unless HR managers are acquainted and employ outcomes of HR analytics this would not be possible.

Many organizations like Google, Wells Fargo, Xerox, 3 M, Ericsson and others have benefitted from HR analytics.

15.7.1 Changing Role of HR Managers

With the emergence of HR analytics, roles of HR managers are also changing, since many of HR jobs can be efficiently performed with the help of HR analytics.

This is helping HR managers to come out of drudgeries of lengthy work processes.

Exhibit 15.9 explains transformation in organizations.

Exhibit 15.7: Transformation in HR Organizations

In recent years, HRM has undergone dynamic changes in its roles, functions and overall impact in shaping organizations. The digital transformation has brought changes in all spheres of HRM. The change basically is seen in the role of Chief Human Resources Officer (CHRO). He/she occupies a very important place in the organization, helping the business to ensure strategic growth. He/she is responsible for decision-making relating to recruitment, compensation, talent management, performance management, training, employee experience and more.

CHROs is now required to offer his/ her inputs on high level business guidance and leadership on defining the very future of work. In this process, organizations are moving from traditional, rigid systems to new agile organizational structures.

Another important change is seen in employee engagement, which is on priority list for CHROs. It is a proven fact that higher the level of engagement of employees in the organization, the higher will be business growth.

The global Human Resource Management (HRM) sector is projected to reach $30 billion by 2025.

Advancements in information technology (IT), predictive analytics, artificial intelligence, and machine learning in HR processes are making companies to invest in technology, but the problem is many of them do not know how to use it optimally. According to a KPMG's (one of the big four accounting organizations) report, 50% of HR leaders are unprepared to use technology available around them. This is the biggest challenge all companies are facing.

HR leaders are in the unique position of leading the future of work, coming out from their circumscribed HR world.

The future of HR has arrived, and it is now. HR thus needs to be a function that takes the lead in understanding what makes people engaged, what cultures drive the most productive workplaces and what equations enable a true merging of human capability with technology.

Source: https://economictimes.indiatimes.com/small-biz/hr-leadership/leadership/the-future-is-now-the-changing-role-of-hr/articleshow/68229542.cms, 2-3-2019

Some of the changes that would be seen in HRM due to using HR analytics are:

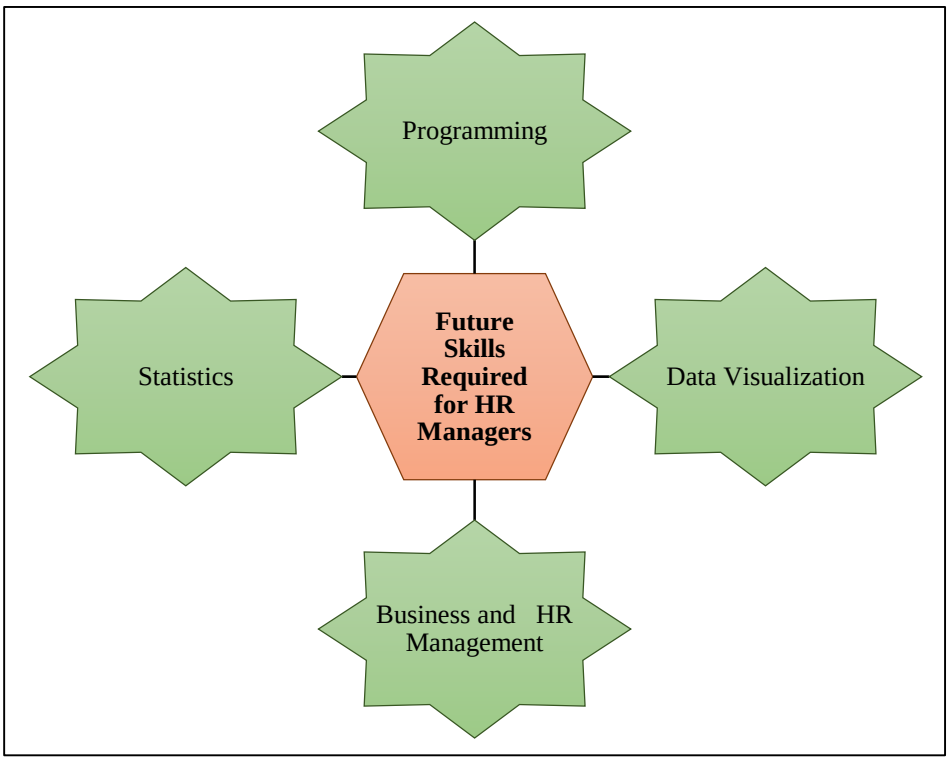
- Matching manpower scheduling based on demand, so as to make available manpower at right time.

- Facing the challenges of talent attrition, talent attraction, development and retention.
- Understanding organizational dynamics and make employee contribution satisfying.
- Future jobs will be talent driven, and make tasks personal and collaborative.
- Providing new insights into competitive advantage making use of data.
- Data driven approach to HRM will be seen in all HR functions.
- Routine HR jobs will be decreasing and more of outsourcing jobs will emerge.
- HR functions will be more inclusive and participative.
- Entry level jobs and routine and transactional task may become redundant.

15.7.2 Future Skills Required for HR Managers

With changing roles of managers in several areas, it is obvious that HR managers require some new skills to cope up with the demands. In the context of increased use of HR analytics, along with predictive analytics by several organizations, we can identify some common generic future skills required for HR managers (see Figure 15.8).

Figure 15.8: Future Skills required for HR Managers



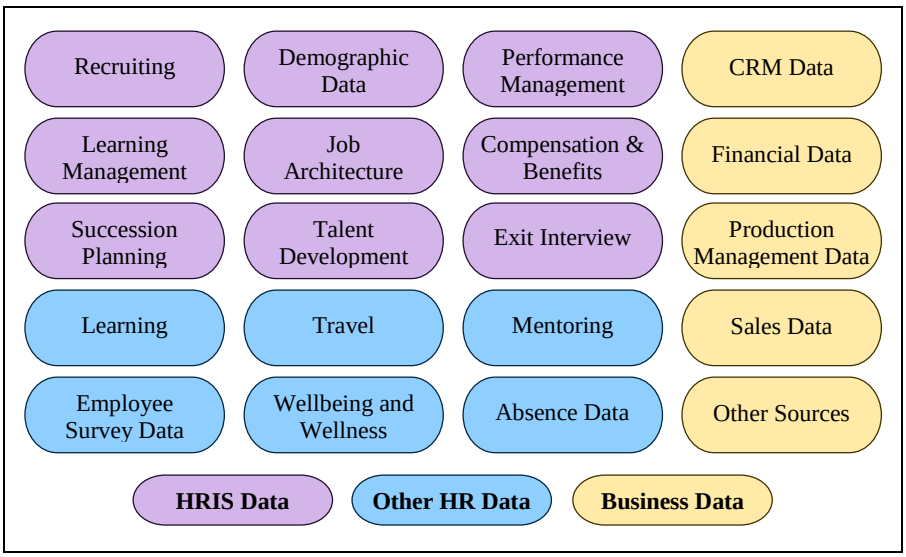
Source: Bhattacharya Dipak Kumar (2017).Understanding Theories and Applications.HR Analytics.

SAGE Publications India Pvt Ltd-May, 2017, ISBN: 9789386062710

- **Business and HR Management:** HR managers should have thorough knowledge of business and human resources for meeting the future requirements. They should be able to understand thoroughly the causes for attrition and techniques to retain talent pool. In addition, they also should be equipped with knowledge on sales, changing needs and requirements of customers, sales margins and sales objectives. They should be able to understand the factors that would increase ROI of the organization.
- **Statistics:** HR analytics requires statistical techniques along with data analysis and data interpretation. When machine learning (MI), artificial intelligence (AI) are being used widely in organizations, it is necessary for HR managers to learn and be ready for meeting the requirements to use these techniques for improvement of HRM. HR managers should be equipped with statistical and mathematical knowledge which are used in HR analytics abundantly, and should not lag behind these skills to utilize this technology to the maximum.

Some popular HR data sources are given in Figure 15.9.

Figure 15.9: HR Data Sources



Adapted from Vulpen Van Erik. 21 HR Data Sources for Analytics.AIHR Blog. https://www.analyticsinhr.com/blog/hr-data-sources

HR managers need to get familiarized with popular software analytic tools, which are discussed below.

- ✓ IBM Kenexa: Talent Management, talent acquisition, learning of staff data, analyzes employee performance, and development, performance management, talent analytics and surveys.
- ✓ Oracle HR Analytics: HR analytics system that provides comprehensive view and suggests ways of encouragement.

- **People Analytics**: Enabling organizations to achieve their best future through advanced and predictive people analytics
  - ✓ iTrent: MHR offer solutions for Visualizing, Intelligent Reporting, Planning and Predicting your data to truly understand your People data.
  - ✓ Fuel50: Fuel50 increases employee engagement and retention with gamified career growth tools for career visibility, leader coaching, and more.

Programming: HR analytics solutions are user-friendly and there is no need to use complex programming skills. Hence, with programming skills, HR managers can use simple programming techniques and analyze data using available HR analytics solutions.

Analytics Software tools are:

- R is the most used HR analytics tool.
- Python is another programming language and can be used interchangeably with R.
- Excel. When we talk about HR analytics tools, we shouldn't forget the basics.
- Power BI.
- Tableau.
- Visier.
- Qlik.
- SPSS.

Data Visualization: Data visualization skills are important and HR managers will be able to understand the right story from data and persuade people to use data in the right perspective. With such skillets like that of programming and data visualization, HR managers would be able to manage workplaces effectively in the future by using data driven analytics.

The various HR areas where data capture, visualization, and analytics benefit HR and the organization are given in Table 15.1.

Table 15.1: HR Areas where Data Capture, Visualization and Analytics Benefit HR and the Organization

The highest influential sourcing channel	The accuracy of job planning capabilities	Compensation or Benefit Revenue Ratio
The time needed for new hires to reach full productivity	The likelihood of employees to recommending company as a good place to work	HR Cost per Employee

Contd….

Candidates interacting with employer brand actually convert	Likely amount of non-productive effort employees are putting in	Recruitment cost per employee
The diversity make-up of pool of recruits	The access to the necessary benefits	Revenue per employee
The money being spent per hire	The last time the employee was recognized	Performance and potential
The retention rate for a specific manager	The status of organizational health	Billable hours per employee
The impact of training after course completion	The participation level for L&D programs	Engagement rating
The revenue earned per employee	Effectiveness of internal hiring mechanism	Cost of HR per employee
The change in productivity after implementing a new tool	The ROI of L&D programs	Absenteeism

Source: ICFAI Research Center

15.7.3 Empowerment of HR Functions through HR Analytics

It is easier for HR managers to become empowered, if they can use HR analytics effectively. HR analytics helps managers to perform their tasks easily, quickly and effectively. This enables HR managers to focus on other important tasks including decision-making. With the emergence of HR analytics, it has become easier for HR managers to align HR activities with business activities. This means that all HR functions are close to business activities.

With HR analytics, there is enormous change in the performance of HR managers. Their contribution to the company results in higher productivity, and employee satisfaction. Predictive analytics of HR is helping managers in driving HR strategies with future perspectives.

These changes are making HR managers empowered especially that of HR predictive analysis, leading to the effectiveness of HR functions.

15.7.4 Ethical Issues in HR Analytics

HR analytics deals with large data of employees regarding their age, experience, pay, performance, skills and other vital information which is of a personal nature.

Not only the basic information of employees but also data on social media interaction, behavioral patterns, etc. Ethics requires that this data on employees should not be misused. This is all the more important for Analytics as it contains huge information. Organizations have to mark a strict line between what is ethical and what is unethical for using employees' data. Organizations have necessarily to inform and take employee's consent in using personal data. Sometimes, personal data may be useful in organizational growth, but ethics demands using personal data only for legitimate purposes and should not be used without the consent of employee

Example: Seeing Machines (An Australian Driver Monitoring Systems Manufacturer) uses AI based Analytics Solution to alert Fatigued Drivers

Seeing Machines is an Australian Company which manufactures driver monitoring systems. The company developed a HR analytics solution using futuristic technology like AI and Machine Learning for alerting fatigued drivers. This has enabled the company to prevent accidents and deaths. The idea is to increase safety for the drivers. This resulted in highly motivated alert drivers and enhanced the productivity.

The system uses a camera, GPS and accelerometer to track eye and eyelid movement (frequency and duration of driver blinks). When a driver closes eyes for more than 1.6 seconds, an alarm is triggered inside the truck – (noise and a vibration in the seat). Then, if the alarm is repeated for a second time, supervisor is alerted to make contact with the driver via radio.

Source : HR Analytics Applications – Human Resources Management – 2nd Ontario Edition (pressbooks.pub) Library article, 2020, Accessed on 28/09/2022

Check Your Progress - 4

10. Which is not among the changes that would be seen in HRM, due to using HR analytics?
    a. Matching manpower scheduling based on demand, so as to make available manpower at right time.
    b. Facing the challenges of talent attrition, talent attraction, development, and retention.
    c. Understanding organizational dynamics and make employee contribution satisfying.
    d. Future jobs will be talent driven, and make tasks personal and collaborative.
    e. Data driven approach to HRM will be seen in some selective HR functions.

15.8 Summary

- Analytics is defined as scientific data manipulation.
- Business analytics is scientific data manipulation for better decision-making in businesses.
- The scientific approach to human resource management in organizations has given birth to HR analytics.
- HR analytics helps enforce positive outcomes of talent management, performance improvement, employee engagement, and so on.
- HR analytics is useful in making better decisions, improving operations, understanding customers, and monetizing data.
- Descriptive HR decision-making process uses HR metrics and Human Resource Information Systems (HRIS), to get insights on decisional issues and on the basis of which decisions are taken.
- Correlation HR decision-making is used to know the relationship between two or more variables.
- Predictive HR decision-making uses predictions and tools such as causation and regression analysis.
- HR functions in the future will have to use analytics for more effectiveness.
- With the emergence of HR analytics, roles of HR managers are also changing and many HR jobs can be efficiently performed with the help of HR analytics.
- With growing importance of HR analytics, HR managers require additional skills like programming, proficiency in statistics, data visualization, and business management.
- As HR analytics helps managers to perform their tasks easily, quickly, and effectively, they are able to concentrate on other vital issues also and are getting empowered.
- HR analytics consists of huge personal data of employees and, hence, it is all the more essential to follow ethical guidelines while using this personal data.

15.9 Glossary

Advanced Analytics Phase: This is the phase of proactive identification of problems or decisional issues for reaching actionable solutions. This is the phase from where we actually begin the use of analytics solutions to solve our business problems.

Analytics: It is defined as scientific data manipulation which can measure and report facts or metrics over a period explaining how variables are related to one another.

Correlational HR Decision-Making Process: It helps in assessing the relationships between two variables.

Descriptive HR Decision-Making Process: It makes use of metrics or HRIS to get insights into decisional issues and then take decisions.

Future HR Analytics: It is integrated with strategic-level business analytics. HR analytics in future would become more technology-enabled and would require specialized skillsets.

HR Analytics: It is defined as the application of analytic logic for HRM function so that it can benefit organizations in improving the performance of employees, help in rationalizing HR decision-making process, and can also improve the ROI from human resources.

HR Decision-Making Process: It is HR managers' judgmental thoughts, encompassing all stakeholders on any action. An effective HR decision-making process requires integration of critical thoughts and information.

HR Metrics: HR metrics are measurements used to determine the value and effectiveness of HR initiatives.

HR Scorecard: HR scorecard is a visual representation of key measures of human resource department's achievements or measurements of important factors of the organization.

HR Forecasting: It focuses on measuring the future requirement of manpower and its implications of human resources on organizational strategy.

Metrics: These are measured in terms of counts, percentages, ratios, and so on.

Operational Phase: This phase involves the task of operational reporting, e.g., performance, compliance matter, and so on. This is also known as the reactive phase.

People Analytics: It is the other way of naming HR analytics. It integrates HR functions with sales, customer retention, accidents, frauds, and quality issues, and then performs measurement of data to get new insights for better decision-making.

Predictive HR Analytics: It blends data to develop an algorithm, based on which HR managers can pre-assess future events as consequences of current HR decisions. It can even help in understanding the behavioral changes of employees.

Time to hire Ratio: Measured in terms of cycle time required from initial HRP to actual recruitment in terms of job placement.

15.10 Self-Assessment Test

1. What is HR analytics?
2. Discuss critical HR decision-making in relation to HR analytics.
3. Explain the future roles required for HR managers.
4. What are the changing aspects in HR with the emergence of HR analytics?

15.11 Suggested Readings/Reference Material

1. Maleh, Yassine. Shojafar, Mohammad. Alazab, Mamoun. Baddi, Youssef. Machine Intelligence and Big Data Analytics for Cybersecurity Applications (Studies in Computational Intelligence, 919) 1st ed. 2021 Edition.
2. Ahmed, Syed Thouheed. Basha, Syed Muzamil. Arumugam, Sanjeev Ram. Patil, Kiran Kumari. Big Data Analytics and Cloud Computing: A Beginner's Guide, 2021.
3. Saleem, Tausifa Jan. Chishti, Mohammad Ahsan. Big Data Analytics for Internet of Things 1st Edition, April 2021.
4. Jones, Herbert. Data Science: The Ultimate Guide to Data Analytics, Data Mining, Data Warehousing, Data Visualization, Regression Analysis, Database Querying, Big Data for Business and Machine Learning for Beginners Hardcover – 10 January 2020.
5. Maheshwari, Anil. Data Analytics Made Accessible: 2023 edition Kindle Edition
6. Mayer-Schönberger, Viktor. Cukier, Kenneth. Big Data: A Revolution That Will Transform How We Live, Work, and Think Paperback – October 26, 2021.

15.12 Answers to Check Your Progress Questions

1. (d) Better Decision-Making

Business analytics makes use of mathematical and statistical techniques for scientific data manipulation for better decision-making.

2. (d) Organizations lie anywhere on the HR analytics spectrum, depending on the maturity of the HR processes, quality of data, and available capabilities

3. (c) Retain intuition or guess work in recruitment.

Guess work or gut feeling does not work often and can cost companies greatly. A data driven approach to recruitment helps companies find employees, who are well suited to the needs of the organization.

4. (e) Just measure some of the terms like rate of absenteeism

Earlier, HR managers just measured some of the terms like rate of absenteeism, attrition, cost of compensation, and the like. These were not sufficient for efficient decision-making needs of the HR manager. With the help of HR analytics, managers can now assess employee engagement, predict the future requirements, and assess the customer relationship management practices.

5. (b) Here data is analyzed to predict future behavior and explains what is likely to happen in future

In Predictive analytics model data is analyzed to predict future behavior. And explains what is likely to happen in future.

6. (e) Measurement and adjustment

The sensitivity and actual outcome of the analytics process is taken care of measurement and adjustment stage. The measurement scale is improved, if found necessary and modified, to suit the ongoing market requirements. This makes HR analytics a cyclic process as feedback is incorporated and scales reworked as per the inputs.

7. (a) Operational Phase

Operational phase is also called reactive phase. This involves the task of operational reporting on performance issues like compliance matters and so on.

8. (c) Advanced analytics phase

Advanced analytics phase is also known as proactive identification of problems or decisional issues for reaching an actionable solution. In this phase, statistical modeling techniques, root cause analysis, and the like are performed for solving business issues.

9. (a) Uses HR metrics and Human Resource Information Systems (HRIS) to get insights on decisional issues

Descriptive HR decision-making process uses HR metrics and Human Resource Information Systems (HRIS), to get insights on decisional issues and on the basis of which decisions are taken.

10. (e) Data driven approach to HRM will be seen in some selective HR functions

Data driven approach to HRM will be seen in some selective HR functions only. It is not among the changes that would be seen in HRM due to using HR analytics.

Unit 16

Data Analytics for Top Management Decision Making

Structure

16.1 Introduction

16.2 Objectives

16.3 Business Intelligence

16.4 Business Analytics

16.5 Correlation Analysis

16.6 Regression Analysis

16.7 Multiple Linear Regression

16.8 Logistic Regression

16.9 Factor Analysis

16.10 Exploratory Factor Analysis (EFA)

16.11 Principal Factor Analysis (PFA)

16.12 Confirmatory Factor Analysis (CFA)

16.13 Classification

16.14 RFM (Recency Frequency Monetary) Analysis

16.15 Market Basket Analysis (MBA)

16.16 Summary

16.17 Glossary

16.18 Self-Assessment Test

16.19 Suggested Readings/Reference Material

16.20 Answers to Check Your Progress Questions

"It is a capital mistake to theorize before one has data."

- Sherlock Holmes in "A Study in Scarlet" by Arthur Conan Doyle

16.1 Introduction

The quote emphasizes the importance of using data to drive smart decisions. Ideas or theories without data are just assumptions, with no factual reality to back them up. If top management truly wants to understand what's going on in the chosen market, they need data.

In the previous unit, we have discussed unstructured data handling methodologies in detail. We have also studied the conversion strategy of unstructured data to structured data which facilitates theenvironment to apply data analytics.

In the present unit, we will discuss data analytics at large to understand the inference procedure out of the big/enormous data collection.

Focus on business analytics has increased in the past decade, reaching many organizations and a wider range of users like executives, business managers, analysts and knowledge workers within the organizations. With fast growing data volumes due to increasing use of applications in companies, business analytics allows to optimize operations and at the same time to maintain flexibility.

There are many statistical models developed and extended to execute and perform business intelligence and business analytics. Various tools like correlation analysis, regression analysis, multiple linear regression and logistic regression are discussed in this unit. data mining techniques like exploratory factor analysis (EFA), principle factor analysis (PFA), confirmatory factor analysis (CFA), classification, predictive analysis, cluster analysis, association analysis, RFM (recency frequency monetary) analysis and market basket analysis (MBA) are also explained in this unit.

16.2 Objectives

After going through this unit, you will be able to:

- Explain business intelligence and how it is useful for organizations
- Describe correlation analysis
- Discuss regression analysis
- Describe multiple linear regressions
- Define logistic regression
- Explain RFM (recency frequency monetary) analysis
- Define market basket analysis (MBA)

16.3 Business Intelligence

Business intelligence (BI) is a broad category of application programs and technologies used for gathering, storing, analyzing and accessing data to help users make better business decisions. BI applications support the activities of query and reporting, decision support, online analytical processing (OLAP) (computer processing that enables a user to easily and selectively extract and view data from different points of view), forecasting, statistical analysis, and data mining (examining large pre-existing databases in order to generate new information). Business intelligence includes concepts/methods and by using fact-based support systems, it improves business decision-making.

Business intelligence is an art of gaining business advantage from data by answering various fundamental questions. These may include, how various customers rank products, organizations, etc., how business is doing at the present stage, and if continued on the current path, which clinical trials should be taken further and which should be stopped.

16.3.1 Business Intelligence: Components

In most cases, business intelligence involves multidimensional analysis and reporting, often based on the company data warehouse to organize the needed data.

Business intelligence includes various key components that are explained in more detail in the following sections:

i) Multidimensional analysis

ii) Reporting

iii) Data mining

iv) Financial consolidation and budgeting

v) Key Performance Indicators

i) Multidimensional analysis

This area covers the possibility to slice-and-dice the data (that is, the factual information) in many dimensions. This is known as pivoting data. A pivot table is a tool to build and summarize data using spreadsheets. In Excel Sheet, one can summarize data in a pivot table mode on many levels on each dimension.

ii) Reporting

Companies need different types of reports. In many cases, hundreds of different types of reports, and often more are needed. Business Intelligence software often has comprehensive reporting tools that can be applied to real-time data available from internal web pages, internet and Excel and PDF format. In many cases, these reporting facilities will be controlled by parameters that can be chosen in real-time

iii) Data mining

Data mining, a branch of computer science, is the process of extracting patterns from large data sets using a combination of statistics and artificial intelligence approaches to study the given data to provide actionable intelligence.

iv) Financial consolidation and budgeting

Business intelligence methodology also covers systems and functionalities for groups to perform financial group consolidation and budgeting. BI tools help in

Reporting, Budgeting and dashboard modules are:

- It creates complete financial statements and operational reports without having to learn proprietary report writers or complex formulas.
- Users can refresh reports on-demand and drill down to detail, eliminating the need to export and reformat reports or the use of multiple report writers across transaction systems. It quickly creates and deploys input templates based on the existing budget model, to completely redesign and modernize the model.
- It can consolidate many data sources into the tool for integrated reporting and analysis.
- Build quick dashboards, highly summarized and Key performance indicator -based goals for executives, and detailed operational dashboards for line managers and end users across the organization.

v) Key Performance Indicators

The key performance indicators are metrics that are measured periodically to keep track of historical trends as well as the goal or target (as per the suitability of conditions).

Example: The New York Shipping Exchange more than Tripled its Shipping Volume from Asia to the United States Due to Business Intelligence and Other Efforts

The New York Shipping Exchange (NYSHEX) is a shipping-technologies company with a mandate to improve the process of exporting goods from the United States.

The company used to manually extract data from its legacy systems and many cloud based applications and import to Excel. The process used to consume time and also only few employees used to have access to the right information and insights to take decisions. Employees have to ask for reports from IT and analytics departments. This was leading to delays and so many decisions could not be based on analytic based insights.

The company invested in Business Intelligence; data was centralized. BI tools were deployed which enabled employees with no coding experience to generate their own reports instantaneously without depending on IT team.

This resulted in the company tripling its shipping volume.

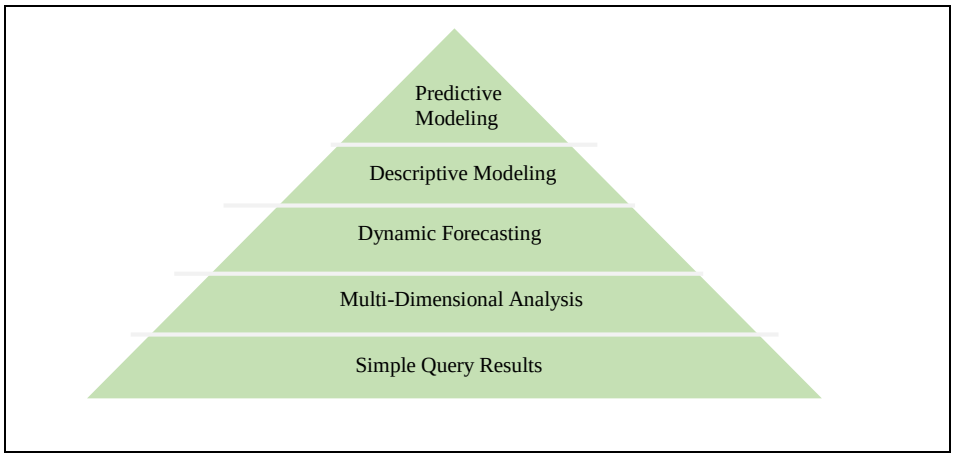
Source: Business intelligence strategies: Examples, techniques, roadmap, and more (2022) - Dataconomy date June 13, 2022, Accessed on 28/09/2022

16.4 Business Analytics

Business Analytics (BA) is a great tool that helps in decision-making. It helps in in increased profitability, reduced cost, faster decision-making, and critical performance in a business.

Business analytics mostly focuses on creating different insights and understanding of the business performance. It uses statistical methods, quantitative analysis, predictive modeling and fact-based management to drive decision-making (see Figure 16.1). For example, the questions answered by Business analytics would be: “Why is this happening?”, “What if these trends continue?”, “What will happen next?”, and “What will be the optimal outcome?”. The latest being that we are moving from diagnostic to prognostic to prescriptive, covering "Where are the recommendations made for optimal approaches to meet the goals?"

Figure 16.1: Business Analytics Pyramid



Source: https://enterprisemanagement.com/johnmyers/business-analytics-2013

The most important characteristics of business analytics for organizations are the use of analytics to get an inside view of data and the facts behind, to implement it in strategic planning and potential decision-making by senior management. This helps them to make better decisions by accessing real-time data which was earlier accessed and used only by IT-aware knowledge workers. Some applications of Business Analytics for businesses to optimize are critical product analysis, up-selling opportunities, improved customer services, better inventory management, and competitive price insights

Example: Netflix deployed Business Analytics for Demand Forecasting and saved Around 1 Billion Dollars Every Year

Netflix has been delivering the content its viewers want, by using big data and machine learning to get insights into viewer preferences, be it how they consume content or what they want to consume.

Contd….

Their "% match" rating is the latest example of how they provide data-based recommendations. This data has facilitated the company to take tactical decisions such as releasing all episodes of a full season together, auto-play the next episode, and offer recommendations for similar or related film or show. Netflix original content was also created by insights provided by Data analytics.

Source: Machine learning in retail use cases & best examples (2022) - Dataconomy date 31/08/2022, Accessed on 29/09/2022

16.5 Correlation Analysis

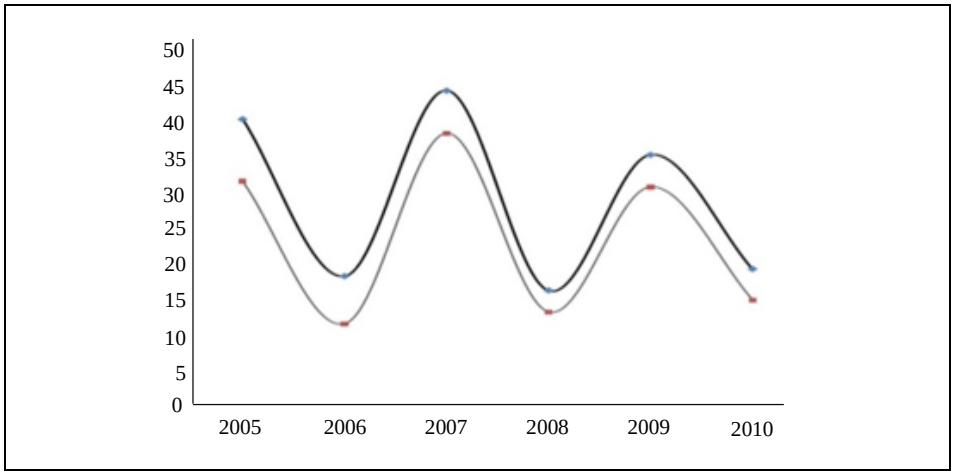
After voluminous data is gathered and stored to be analyzed for making better decisions, it is very much essential to identify if any relationship exists between the variables present in the data. This not only facilitates organizations to take better decisions, but also assists them to build an association between the variables.

Statistics gives an analytical approach in similar situations. Statistics helps in predictive analysis from the existing data. Thus it is quite useful in business intelligence.

Correlation Analysis is used as a statistical tool to discover the association between bivariate variables. It may be noted that correlation analysis is one of the most widely used statistical techniques adopted by the statisticians to find the relationship between the variables.

Many times we come across problems or situations where two variables seem to move in the same direction, either both increasing or both decreasing. At times, an increase in one variable is accompanied by a decline in another. Such changes in variables suggest that there is a certain relationship between them (see Figure 16.2).

Figure 16.2: An Example of Graphic Correlation



Source: ICFAI Research Center

When we use correlation analysis and establish a relationship between two variables, then we confront a major question: Does this relationship indicate the existence of cause and effect relationship? A multidimensional analysis is done to know which variable is having more influence over the other.

The correlation may exist by chance, particularly when a small sample of data is involved. Also for a small sample series, no relationship may exist. It is possible that both the variables are influenced by one or more other variables or there may be another situation where both the variables may be influencing each other. This makes it difficult to establish the cause and the effect. The foregoing discussion clearly shows that correlation does not indicate causal and functional relationships. Even when there is no cause-and-effect relationship in bivariate series, if one interprets the relationship as causal, such a correlation is a spurious correlation.

Based on the data available in the organization, correlation analysis can be performed on various variables. Correlation can be positive or negative, linear or non-linear or simple, partial or multiple. The correlation coefficient closer to 1.0 indicates a strong relationship and less than 0.5 indicates a weak relationship between the two variables. The range is [-1,1]. Technically closer to -1 is also a strong correlation, but a negative correlation.

In business organizations, correlation analysis can be extremely helpful. It has been used extensively in agriculture, economics, business and several other fields.

It can be enabled to estimate costs, sale prices and other variables on the basis of some other variables inferring closeness of the relationship with variables concerned. When a specific and reliable relationship has been established between any two given variables, we can find the value of a variable given the value of another. In fact, this is done with the help of regression analysis, which is discussed in the next section (11.6); however, regression analysis is a predictive tool that predicts the future state of the relationship between the given variables whereas correlation analysis is the indicator of the current state of the relationship between the variables taken. This also shows that the two concepts, correlation and regression analyses, are closely related with a thin line of segregation between them.

In business, sometimes forecast is necessary in order to take a decision regarding a product or a particular course of action. To forecast, some relationship between a pair or group of variables relevant to a particular situation are to be ascertained. For example, a company wants to know how the sales will increase in the next five years, along with the growth of population and increase in demand of the product. Here, it is assumed that the increase in population will lead to an increase in sales. Thus, it is important for the company to determine the nature and extent of the relationship between these two variables.

Example: Supplyframe Uses Correlation Analysis and Other Techniques to Provide the most Comprehensive Metal Price Intelligence Solution to its Clients

Supplyframe is the intelligence platform for the global electronics value chain. The company deploys AI and "correlation analysis" based solutions to sense and interpret billions of 'intent, demand, supply, and risk signals 'to deliver insights at key decision points throughout the entire design-to-market product lifecycle. Over 10 million engineering and supply chain professionals worldwide engage with the company solutions to power innovation and optimize over in excess of $120 billion in annual direct materials spend. The company is assisting its customers and the market to understand metals and other raw material dynamics to predict and normalize technology supply chain risk. The customers will get valuable insights into the supply constraints and pricing pressures leading to global volatility in metals.

Source: Supplyframe Commodity IQ Expands Raw Materials Intelligence with MetalMiner | Business Wire date 05/08/2022, Accessed on 29/09/2022

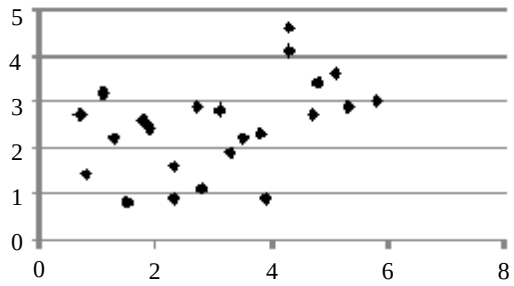
16.6 Regression Analysis

Regression model is defined as "a statistical model with a set of mathematical formulae and assumptions which describe a real-world situation". A statistical model tries to capture the systematic behavior of the given data, leaving out those factors that cannot be foreseen or predicted. Despite our best efforts, it is highly unlikely that a model may reveal a perfect real-world situation.

A good statistical model is one which provides as large a systematic component as possible, minimizing errors. These errors are on account of a number of factors that we are unable to identify. In case we are able to construct a good model, then the average of observed errors will be zero. These errors should also be independent of one another.

A common example in the business world is the relationship between advertising and sales. When a linear regression model involving these two variables is appropriate for prediction, we may use it for predicting sales for a given level of advertising expenditure. It may be noted that the level of advertising should be within the range of expenditure on the advertising covered. A scatter diagram can give us a broad idea of the type of relationship (or even absence of any relationship) between the two variables. When plotted on the scatter diagram, it visually gives a feeling of the type of relation between the two variables on the graph as shown in Figure 16.3.

Figure 16.3: An Example of Scatter Diagram



Source: ICFAI Research Center

While regression analysis is an extremely useful technique for making predictions and as such frequently used, one should be careful in avoiding errors that may arise on account of the wrong application of regression analysis.

Example: Rajiv Gandhi Super Speciality Hospital at Delhi uses Regression Analysis to Find that Hypertension is the most Common Morbidity in COVID-19 Patients

Rajiv Gandhi Super specialty Hospital is a government run hospital at Delhi. The doctors of the hospital wanted to study the epidemiological characteristics of the pandemic, based on the clinical profiles of COVID-19 patients. The study used "Regression Analysis" technique. The insight obtained from the study is that hypertension was the most common co-morbidity among the Covid 19 patients. Regression analysis was used and a P value (correlation coefficient) less than 0.05 was statistically significant. The study results will help public health authorities to put in policy interventions.

Source: Hypertension most common co-morbidity: Delhi hospital study on Covid patients | The Financial Express date 25/05/2022, Accessed on 29/09/2022

16.7 Multiple Linear Regression

In the preceding sections, the discussion was confined to only two variables. However, in business, we come across situations where the relationship is not that simple. One variable may be affected by two or more independent variables. For example, the sale of a product may be related to a number of independent variables such as price, income, advertising expenditure, seasons, number, size and location of retail outlets, quality of the product and so forth. If in such cases, we take cognizance of only one independent variable, then the magnitude of error in the result is likely to be high. In the light of this, it is desirable to use two or more independent variables in the estimating equation. The statistical technique of extending linear regression methods, so as to consider two or more independent variables is known as multiple linear regression. Multiple regression, as a

predictive analysis, is used to explain the relationship between one continuous dependent variable and two or more independent variables. The independent variables can be continuous or categorical (dummy coded as appropriate).

In multiple regressions, many formulae can be used to ascertain the relationships among variables taken. Tedious calculations are involved in multiple regression analysis. To overcome the problem, computers and other software applications are used. This facilitates us enormously as several independent variables can be handled. We also can ascertain whether adding another independent variable will improve our results or not.

Multiple regression analysis is useful in as much as it shows the degrees of association between one variable taken as a dependent variable while the remaining variables, two or more, are taken as the independent variables. It also serves as a measure of goodness of fit for a given series of data.

Constraints: A major limitation of multiple regression analysis is that it assumes that the relationship amongst the variables is linear. However, we find in practice a large number of relationships that are not linear and follow some other pattern. Another constraint is based on the assumption that the effects of independent variables on the dependent variable are quite separate from each other and hence, additive. Also, the amount of work involved in the calculation of multiple linear regressions is enormous.

Activity 16.1

Data for Business Forecasting

A company involved in selling the garments has hired you for giving the forecast of procurement of stock to supply against the demand. What type of data will be required for the exercise and which type of analysis you would be doing to decide on the forecast?

Answer:

16.8 Logistic Regression

The crucial limitation of linear regression is that it cannot deal with discrete variables that are dichotomous and categorical. Many interesting variables in the business are categorical in nature, for example, consumers making a decision to buy or not to buy a product may pass or fail the quality control, etc. A range of regression techniques has been developed for analyzing data with categorical dependent variables, including logistic regression.

Since the dependent variable is dichotomous, one cannot predict a numerical value for it using logistic regression. So, regression based on 'least squares deviations' criteria (fitting curves with least error square from original data to the new analytical fitted data). This is the best fit approach of minimizing error around the line.. Instead, the binomial probability theory based logistic regression (Logistic regression is a technique for making predictions when the dependent variable is a dichotomy, and the independent variables are continuous and/or discrete.) is used, in which there are only two values to predict, i.e., the event/person belongs to either one group or the other. Logistic regression forms a function, based on using the maximum likelihood method (Maximum likelihood, also called the maximum likelihood method, is the procedure of finding the value of one or more parameters for a given statistic which makes the *known* likelihood distribution a maximum), which maximizes the probability of classifying the observed data into the appropriate category given the regression coefficients.

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between one or more categorical or continuous predictor variables and the categorical outcome variable.

There are two main uses of logistic regression:

1. First is the group membership prediction problem which involves predicting whether or not a collection of instances share a certain semantic property. For instance, in a verification given a collection of images, the goal is to predict whether or not they share a {\it familial} relationship. Since logistic regression computes the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio.
2. Logistic regression also provides knowledge and strengths of the relationships among the variables.

Many procedures in SAS/STAT like CATMOD, GENMOD, LOGISTIC, and PROBIT can be used to perform logistic regression analysis. Every procedure has a special feature that makes it useful for certain applications.

Check Your Progress - 1

1. Data mining is also referred to as which of the following?
    a. Knowledge Discovery in Databases
    b. Data Cleaning
    c. Data Extraction
    d. Data Management
    e. Data Mart

2. An OLAP tool is provided for which of the following?
    a. Multidimensional analysis
    b. Slicing and dicing
    c. Roll-up and drill-down
    d. Rotation
    e. Setting up only relations
3. Which of the following is not used by business analytics?
    a. Statistical tools
    b. Quantitative techniques
    c. Predictive modeling
    d. Operations research modeling
    e. Median analysis
4. If two variables are highly correlated, what can you infer?
    a. They always go together.
    b. High values on one variable lead to low values on the other variable.
    c. There are no other variables responsible for the relationship.
    d. You cannot make any of the casual claims, nor can you be sure they always go together.
    e. Both the variables take the same values.
5. In regression analysis, which of the following variables is being predicted?
    a. Response or dependent variable
    b. Independent variable
    c. Intervening variable
    d. Usually the variable itself
    e. Neither dependent nor independent

16.9 Factor Analysis

The process of inspecting, cleaning, transforming and modeling data for discovering useful information that helps to arrive at certain conclusions and support the decision-making process is called data analysis. There are multiple approaches with different techniques for data analysis. The data analysis in statistics is divided into descriptive statistics- used to describe the basic features of the data in a study, Exploratory Data Analysis (EDA) referred as Exploratory Factor analysis (looking for clues in data), and Confirmatory Data Analysis (CDA) - evaluate evidence using traditional statistical tools such as significance, inference, and confidence.

Factor analysis is a multivariate statistical procedure that has many uses. Firstly, factor analysis cuts down a large number of variables into a smaller set of variables (also referred to as factors). Secondly, it establishes fundamental dimensions between measured variables and hidden concepts, thereby allowing the formation and improvement of theory. Thirdly, it provides construct validity evidence of self-reporting scales.

If you are a banker and you are looking at increasing the reach and market share of your bank, you can do so by carefully analyzing the consumer perceptions about banks and their expectations regarding the same. Factor analysis is an important technique of doing that and it involves the effective utilization of the behavioral patterns of different consumers as well as their demographics. India being a vast and culturally diversified country, factor analysis can be much more effectively used here.

Factor analysis is considered to be better than other statistical tools that are available in the market today for analysis purpose owing to the suitability of conditions prevailing.

Factor analysis has certain strengths as well as shortcomings:

1. Factor analysis has a high degree of replicability, which means that the experiment can yield the same kind of results even in different environments.
2. Lot of underlying factors which cannot be explicitly brought out through various statistical analyses can be achieved through factor analysis.
3. If a research is quantitative, then it has lots of components that require subjective interpretation; factor analysis is the best in doing it.

Software like SPSS using Excel and Crosstabs can be very effective in doing factor analysis. The effectiveness is clearly visible in the case of tests involving quantitative research and quantitative data interpretation in various industries. It can also be used in the field data analysis involving research in specialized areas, and also in psychological studies, like intelligence, attitude, behavior, etc. Apart from the above, factor analysis is also used in market research projects in various fields like marketing and sales.

16.10 Exploratory Factor Analysis (EFA)

Factor analysis is a popular collection of heuristic techniques (self-learning) used by analysts as a part of behavioral science. Exploratory factor analysis is a primary technique for many researchers to conduct assessment-related studies. The goal of EFA is to maximize the amount of variance explained, by identifying factors based on data. The researchers need to have specific hypotheses about how many factors will emerge, and what variables these factors will be made up of.

Factor analysis is further composed of two subsets namely “Common variance” and “Specific variance”. Common variance refers to variance attached to

measurement method. Specific variance is attributed to characteristics of various individual indicators. Unrestricted measurement models are estimated in EFA. There is no unique set of statistical estimates for unrestricted measurement models. This property relates to the rotation phase and is part of many applications of EFA. It is assumed in EFA that the specific variance of the individual indicator is not shared between them. Procedures for EFA are available in SPSS and SAS/STAT.

16.11 Principal Factor Analysis (PFA)

Principal factor analysis (PFA), also referred to as the Principal axis factoring (PAF) and common factor analysis, aims to identify the minimum number of factors that can lead to correlation between a given set of variables, whereas the more common principal components analysis (PCA), in its full form, seeks the set of factors which can account for all the common and unique (specific plus error) variance in a set of variables. PFA is generally used when the research purpose is detecting data structure (latent constructs or factors) or causal modeling.

16.12 Confirmatory Factor Analysis (CFA)

Like exploratory analysis, confirmatory factor analysis (CFA) is also widely used in statistical analysis. CFA is used to test a proposed theory or model and unlike EFA, it has assumptions and expectations based on priori theory regarding the number of factors, and which factor theories or models are the best fit. The major advantage of CFA is to study the relationships between a set of observed/continuous latent variables.

Confirmatory factor analysis (CFA) is a restricted measurement model. That is, the researcher must explicitly specify the indicator-factor correspondence to evaluate CFA. The model considers multivariate regression to describe the relationship between a set of dependent variables and latent variables. The observed dependent variables are referred to as factor indicators and the continuous latent variables are referred to as factors. CFA is a method to specify which variables load onto which factors. Based on the goodness-of-fit of the defined model, the result is taken, or modifications are made to the originally defined structure.

16.13 Classification

Of late, a large amount of data is being collected and maintained in databases across the business world.

There is a lot of information and knowledge that can be extracted from such databases; and with automation for extracting this information, it is possible to mine the data. There are different methodologies to tackle such problems, such as classification, association rule mining, clustering etc.

Classification divides customer records into distinct segments referred to as classes. Classification analysis requires that the end-user/analyst knows how classes are defined. The objective of a classifier is to decide how new records should be classified, for instance, "is a new customer likely to default on the loan?"

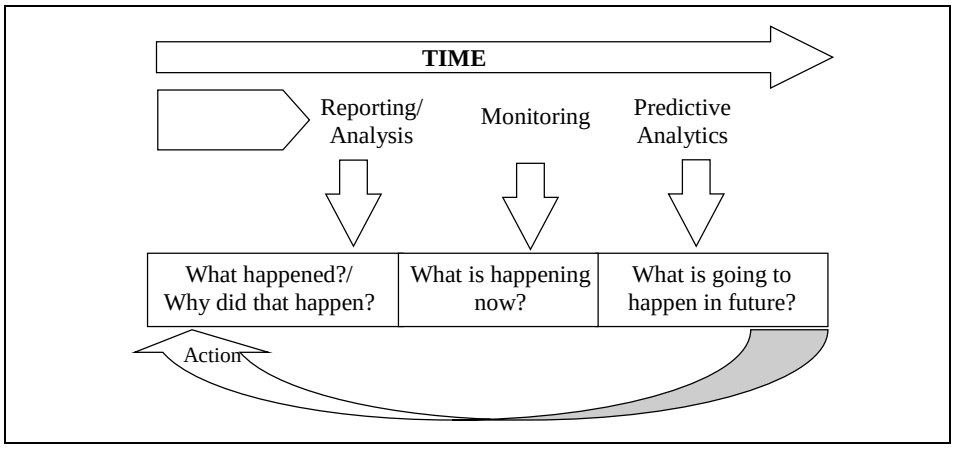
Classifiers use approaches such as decision trees to partition and segment records. New records can be classified by traversing the given tree from its root through branches and nodes, to a leaf representing a class. The path any record takes through a decision tree represents a rule. For example, If "income<Rs.30,000 and age<25, and debt=High", then Default Class=Yes.

16.13.1 Predictive Analysis

The organizations today need to know what is happening to their business, and also be able to predict what is likely to happen. The greatest challenge in the industry is sustainability in the market with persistent growth. Strategic planning has begun to play a heavy role in companies to decide on the future evolution.

The form of the predictive model varies, depending upon the behavior or event that is being predicted. Most predictive models generate a score called the "credit score." The higher the score is, the more likelihood of the event occurrence. Predictive analytics makes use of a variety of statistical and analytical techniques to build models predicting future events or behaviors as shown in Figure 16.4.

Figure 16.4: The Process of Predictive Analytics



Source: Imanuel (2016).What is Predictive Modeling?APM Strategy Whitepaper.

PAT Research-2016
https://www.predictiveanalyticstoday.com/predictive-modeling

Predictive analytics is derived from the data mining model and focuses on predicting future possibilities and trends. Predictive analytics, along with predictive models and data mining techniques, depends on high-end statistical methods, which include multivariate analysis techniques such as advanced regression and time-series models. The insurance industry has always relied on

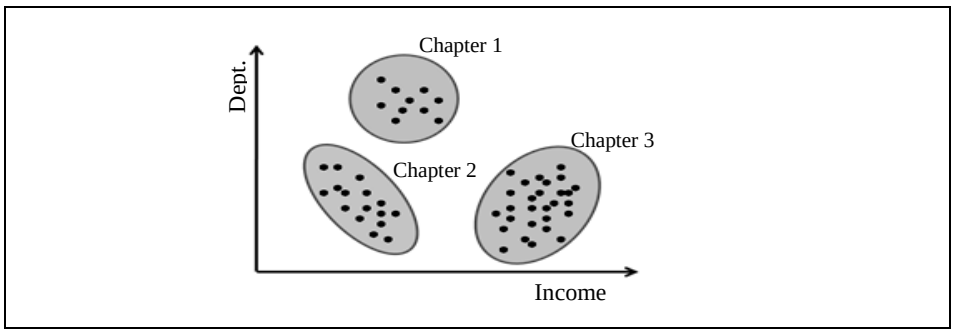
forecasting. The use of predictive analytics has, therefore, quickly become the industry's best practice. Insurers use predictive analytics techniques to focus on potential clients and to identify potential fraudulent claims.

The applications of predictive analytics are spread over many sectors like CRM, healthcare, cross-selling, fraud detection, risk management, telecommunications and travel.

16.13.2 Cluster Analysis

Researchers in many areas are working on how to organize observed data into meaningful structures, i.e. categorization. Clustering is the tool to group a set of objects in a way that objects in the same cluster group are more similar to one another than to those in the other groups (clusters) as shown in Figure 16.5.

Figure 16.5: A Sample of Cluster Analysis



Source: ICFAI Research Center

Cluster Analysis is a technique of data analysis, where data or information is broken down or clustered into manageable sizes so that interpretation becomes easier. This method of data segregation makes it easy to comprehend and resolve the data issues. Cluster analysis helps in understanding the homogeneity of the data universe (like markets), and the extent of heterogeneity that exists in a datum (market). This can be effectively used in retail, pharmacy, tourism and healthcare sectors.

For example, the tourism sector uses cluster analysis for the following reasons:

1. To identify the homogeneity of the market in a vast country like India, where different heterogeneous income groups exist.
2. To package a brand communication to be delivered to each segment, so that the communication is clear, specific and unambiguous.

Hierarchical Cluster Analysis is a statistical method to find relatively homogeneous clusters based on the measured characteristics. It starts with each case as a separate cluster, i.e., there are as many clusters as there are cases. Then it combines the clusters sequentially. This reduces the number of clusters at each step. The process is done until only one cluster is left. The other methods of cluster analysis are *k*-means clustering, two-step clustering and Ward's method.

16.13.3 Association Analysis

Using association analysis, one can identify groups of customers with similar interests and buying similar products. Using this information, recommendations can be developed that customers who purchased some book of interest also purchased other related books. This can frequently be seen on Amazon interface, where after the initial order for a book or any other article, the recommendation of similar items will start flowing in.

Association analysis identifies relationships or correlations between observations and/or between variables in the datasets. These relationships are then expressed as a collection of so-called association rules. The approach has been proved very successful in mining very large transactional databases like shopping baskets and online customer purchases. Association analysis is one of the core techniques of data mining.

For the online book selling example, historical data is used to identify that customers who purchased two particular books also intended to purchase another particular book. The historical data might indicate that the first two books are purchased by only 0.5% of all customers. But 70% of these customers also purchase the third book. This is an interesting group of customers. As a business, we must take advantage of this observation by targeting advertising of the third book to those customers who have purchased both of the other similar books.

Association rules assist in marketing, targeted advertising, floor planning, inventory control, churning management, etc. In data mining, association rules are useful for analyzing and predicting customer's behavior.

Example: Pratt & Whitney (American Aerospace Manufacturer) used Predictive Analysis to Plan Maintenance of Aircraft Engines

Pratt & Whitney is an American aerospace manufacturer with global service operations. The company is using Collins' GlobalConnect^{SM} solution and Ascentia® Analytics full-flight data product to get aircraft engine data in real time for analysis. The digital service offered by the company analyses data from an assortment of sensors on the engine. It is around 4 million data points per flight. These data points can be used by the maintenance teams to determine well in advance when there will be need to fix or replace a part. This will lead to more effective and proactive maintenance of engines.

Also predictive analysis facilitates better planning; operators can even proactively schedule those repairs in the best locations like an airport with a hangar and many mechanics rather than a remote port with limited resources.

Source: Collins Aerospace Bring its Vision for Connected Aviation to Life (connectedaviationtoday.com) date 28/09/2022, Accessed on 30/09/2022

Check Your Progress - 2

6. Which of the following describes variables used in factor analysis?
   a. Measured at nominal level
   b. Abstract concepts
   c. Not related to each other
   d. Related to each other
   e. Standardized
7. What is the missing element in the list given here for data analysis: cleaning, transforming, and modeling data?
   a. Inspecting
   b. Collating
   c. Extracting
   d. Loading
   e. Transformation
8. If you have multiple predictor variables and a dichotomous dependent variable, then what is the most appropriate multivariate test?
   a. Stepwise regression
   b. Canonical correlation
   c. Logistic regression
   d. Factor analysis
   e. Predictive analysis
9. What is predictive analytics?
   a. Research aimed at anticipating the likely outcome of a course of action.
   b. Designed to generate insights into cause-and-effect relationships.
   c. Research that attempts to provide information on what exists.
   d. Designed to find out what happened in the past.
   e. Nothing related to decision making.

16.14 RFM (Recency Frequency Monetary) Analysis

RFM stands for recency, frequency and monetary Value. It is a database-driven marketing technique that has been used by catalogers to increase conversion rates and reduce the expensive cost of mailing catalogs. Online retailers use RFM analysis to increase conversion rates, personalization and revenue. RMF provides answers to a number of business questions like:

- Can organizations identify their best customers?
- Do companies know who their worst customers are?

- Do companies know which customers they have lost, and which customer they are about to lose?
- Can companies identify loyal customers who buy often, but spend very little?
- Can companies target customers who are willing to spend the most at their store?

16.14.1 How Does RFM Analysis Work?

The goal of RFM Analysis is to divide customers based on buying behavior. One needs to understand the historical actions of individual customers for each RFM factor. Customers can be ranked based on each individual RFM factor, and finally, all these factors together are used collectively to create RFM segments for targeted marketing.

The terms in RFM are:

1. R – Recency
2. F – Frequency
3. M – Monetary Value

1. R - Represents recency of the last purchase. This gives the interval between the time that the last consuming behavior happens and the present one that has taken place. The shorter the interval, the bigger the R.
2. F - Represents frequency, which refers to the number of transactions in a particular period, for example, five times in a year, five times in one quarter or five times in one month. The more the number of times the customer purchases in a given limited period of time, the bigger the F.
3. M - Represents monetary, which refers to the consumption of an amount of money in a particular period. The more the monetary value, the bigger the M.

It was observed that the bigger the value of Recency (R) and Frequency (F) are, the more likely the related customers are to produce a new business with enterprises. Furthermore, the bigger the monetary value (M), the more chances customers buy products or services with the same enterprise again. RFM analysis supports the Pareto axiom: “80% of business comes from 20% of your regular customers”.

RFM has become an important tool that customers are consigned with a ranking of 1, 2, 3, 4, or 5 (5 being the highest) for each parameter in RFM. The three scores are together referred to as an RFM unit. Later, while doing the analysis, the database is sorted to determine which customers are having the unit ranking of ‘555’ and are concluded as the ideal customers.

The limitation of the RFM analysis tool is that the company must be cautious while giving the ranking to the customers. They should also consider that the

customers with low ranking should not be neglected, but instead should be nurtured to become improved customers.

Example: Tata Cliq uses "Recency Frequency Monetary" (RFM Analysis) to Segment Its Customers and Choose Strategies to Retain Customers

Tata Cliq had used "RFM analysis" to segment its customers into 10 segments. This is based on the frequency and how recently the customer opened the app. Since customer retention is one of its core metrics, the company had used RFM to get insights into the segments of customers the company should not lose. The company uses these insights to devise its customer retention strategies.

The company can now easily identify and reward its loyal customers and also identify customers who have not bought for a long time. RFM enabled the company to devise the right messaging strategy for each of these 10 different segments

Source: Clevertap: The Tata Cliq Luxury Story (cnbctv18.com) date 15/04/2022, Accessed on 29/09/2022

16.15 Market Basket Analysis (MBA)

Today, many companies are trying to improve business performance with faster, better decision-making by applying advanced predictive modeling techniques to their huge and growing volumes of data. Business analytics helps in areas like marketing, CRM and operations with valuable insights drawn from their data.

Market basket analysis is a data modeling technique used to find associations between items by determining the likelihood for them to occur together. It is the concept of identifying associations between products that the customers are putting into their shopping baskets. MBA (market basket analysis) is also popularly known as Product Affinity Analysis (PAA) or Association Rule Learning (ARL).

Market Basket Analysis (MBA) is one of the advanced models to leverage voluminous amounts of customer data to determine products, which are most commonly purchased together. Understanding customer's purchasing patterns helps marketing and sales organizations to make more informed decisions about how to deploy their efforts and resources. One of the classic examples of MBA is found in Amazon portal which shows "Customers who bought a specific item also bought allied items A, B and C".

16.15.1 MBA: Understanding Customer Purchase Behavior

Market basket analysis (MBA) is a data mining technique which is widely used in the consumer packaged goods and looks at purchase coincidence. It studies whether any two products are purchased together, and also whether the purchase of one product increases the likelihood to purchase the other.

Market basket analysis helps understanding customers and their purchasing behaviors by allowing companies to explore product associations. It helps in predicting the likelihood of a customer's subsequent purchase behavior based on the associations. MBA is an advanced business analytics tool that can help companies optimize marketing and sales operations for improved performance.

Big data is the collection of a voluminous real data over a period which is subsequently used for many analytics, one of its dimensions being business intelligence.

Example: A Market Basket Analysis of Amazon Fresh in US Finds the Positioning Strategy of the Company

Amazon Fresh is a brick-and-mortar retail format started by Amazon in some locations in the US. The consumer research company Brick-Meets-Click has conducted a study based on the "market basket analysis" approach to find out what is the pricing and positioning strategy of the newly started Amazon Fresh in Schaumburg, Illinois. They have selected a basket of 30 groceries commonly bought by consumers and analysed the prices and sales at Amazon Fresh store and the nearby stores of Aldi, Walmart, and Jewel stores. They tracked the data for six months. The data pertained to daily prices and promotional prices for the 30 items in the basket.

The study found Amazon Fresh positioned itself not as a low-price leader. Its strategy matched with that of high/low grocers like Jewel who start with high price for a product initially and offer discounts as the demand goes down for that product.

Source: What grocers can learn from a market basket analysis of Amazon Fresh | Supermarket News September 8, 2021, Accessed on 29/09/2022

Activity 16.2

Marker Research Process

You are employed by a mobile manufacturing company to give an analysis of the purchase pattern of customers of their product line. It helps in launching a new product in the market. Which type of analysis will be used and how you will go ahead with this process?

Answer:

Check Your Progress - 3

10. In RFM analysis, F stands for which of the following?
    a. Factor
    b. Frequency
    c. Fraudulent
    d. Format
    e. Fiscal

16.16 Summary

- Business Intelligence (BI) has become an expected business competency to improve decision-making effectiveness. It is for all workers, managers and executives to take the most effective action in a given business situation.
- Focus on business analytics has increased steadily over the past decade as evidenced by the continuously growing business analytics software market.
- Business analytics is reaching more organizations and extends to a wider range of users, from executives and number of business managers to analysts and other knowledge workers, within the organizations.
- While the main concern of database technologists was to find efficient ways of storing, retrieving and manipulating data, the main concern of the machine-learning community was to develop techniques for decoding and grasping knowledge from data.
- Many statistical tools were adopted like Correlation analysis, Regression analysis and Logistic regression to find the relationship between the decision variables.
- Other statistical methods are used in data mining for finding the underlying relationships and structures among a large set of variables. Factor Analysis, Exploratory Factor Analysis, Confirmatory Factor Analysis, Predictive Analytics, Cluster Analysis, Association Analysis, Market Basket Analysis, etc., are analytical techniques used in many applications to find the connection between variables.

16.17 Glossary

Bivariate Variable: Bivariate data has two variables and involves relationships between the two variables.

Categorical Variable : Categorical variable is a variable that can take on one of a limited, and usually fixed number of possible values, assigning each individual or other unit of observation to a particular group or nominal category on the basis of some qualitative property.

Continuous Predictor Variable: A continuous predictor variable is a continuous variable used in regression to predict another variable.

Correlation: Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

Data Mining: Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cut costs or both. It allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Dichotomous Variable: A dichotomous variable is one that takes on one of only two possible values when observed or measured.

Discrete Variables: Variables that can only take on a finite number of values are called discrete variables.

Factor Analysis: Factor analysis is a type of statistical procedure that is conducted to identify clusters or groups of related items (called factors) on a test. Factor analysis is used to analyze large numbers of dependent variables to detect certain aspects of the independent variables (called factors) affecting those dependent variables, without directly analyzing the independent variables.

Linear Regression Model: In simple linear regression, a single independent variable is used to predict the value of a dependent variable.

MOLAP: Multidimensional Online Analytical Processing is a kind of Online Analytical Processing (OLAP) that uses a multidimensional data model to analyze data.

OLAP: OLAP (Online Analytical Processing) is computer processing that enables a user to easily and selectively extract and view data from different points of view. OLAP allows users to analyze database information from multiple database systems at one time.

Regression: A statistical measure that attempts to determine the strength of the relationship between one dependent variable and a series of other changing variables (known as independent variables). In a cause and effect relationship, the independent variable is the cause, and the dependent variable is the effect.

Time Series Models: Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

16.18 Self-Assessment Test

1. What is Business Intelligence? What are its components?
2. What are the basic requirements for implementing business analytics? Explain.
3. How is cluster analysis different from association analysis? Explain.
4. Correlation and linear regression are the most commonly used techniques for investigating the relationship between two quantitative variables. What is the core difference between both the techniques?
5. Write short notes on the application of data mining in financial analysis.
6. Explain how Market Basket Analysis can be applied in understanding credit card purchases.

16.19 Suggested Readings/Reference Material

1. Maleh, Yassine. Shojafar, Mohammad. Alazab, Mamoun. Baddi, Youssef. Machine Intelligence and Big Data Analytics for Cybersecurity Applications (Studies in Computational Intelligence, 919) 1st ed. 2021 Edition.
2. Ahmed, Syed Thouheed. Basha, Syed Muzamil. Arumugam, Sanjeev Ram. Patil, Kiran Kumari. Big Data Analytics and Cloud Computing: A Beginner's Guide, 2021.
3. Saleem, Tausifa Jan. Chishti, Mohammad Ahsan. Big Data Analytics for Internet of Things 1st Edition, April 2021.
4. Jones, Herbert. Data Science: The Ultimate Guide to Data Analytics, Data Mining, Data Warehousing, Data Visualization, Regression Analysis, Database Querying, Big Data for Business and Machine Learning for Beginners Hardcover – 10 January 2020.
5. Maheshwari, Anil. Data Analytics Made Accessible: 2023 edition Kindle Edition
6. Mayer-Schönberger, Viktor. Cukier, Kenneth. Big Data: A Revolution That Will Transform How We Live, Work, and Think Paperback – October 26, 2021.

16.20 Answers to Check Your Progress Questions

1. (a) Knowledge Discovery in Databases

Data mining extracts the data patterns and derives knowledge from large databases using different analytical and statistical techniques for better organizational decision-making.

2. (b) Slicing and dicing

An OLAP (Online Analytical Processing) tool provides for slicing and dicing of the database.

3. (d) Operations Research Modeling

Business analytics makes extensive use of data, statistical and quantitative analysis, explanatory and predictive modeling, and fact-based management to drive decision-making.

4. (d) You cannot make any of the casual claims, nor can you be sure they always go together

Correlation does not indicate any causal and functional relationship.

5. (a) Response or Dependent Variable

In regression analysis, the variable that is being predicted is usually the dependent variable.

6. (d) Related to each other

Factor analysis is often used to determine a linear relationship between variables.

7. (a) Inspecting

The process of inspecting, cleaning, transforming, and modeling data for discovering useful information that helps to arrive at certain conclusions and support the decision-making process is called data analysis.

8. (c) Logistic regression

Logistic Regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

9. (a) Research aimed at anticipating the likely outcome of a course of action

Predictive analytics focuses on predicting future possibilities and trends.

10. (b) Frequency

RFM stands for Recency, Frequency and Monitory Value.

Unit 17

Business and Marketing Intelligence Using Analytics

Structure

17.1 Introduction

17.2 Objectives

17.3 Need for Business Intelligence

17.4 Data, Information, Knowledge and Wisdom

17.5 Data Warehousing

17.6 Business Intelligence Components

17.7 Business Intelligence Architecture

17.8 Business Intelligence Methodologies

17.9 Data Mining Techniques

17.10 Market Intelligence and Decision Making

17.11 Making the Last Mile in Data Analytics

17.12 Correlation Analysis

17.13 Market Intelligence Using Analytics

17.14 Customer Experience Management Using Analytics

17.15 Business Intelligence Tools

17.16 Moving Beyond the Tools to Analysis Applications

17.17 Introduction to Google Big Query, Google Dataflow and Apache Spark

17.18 Summary

17.19 Glossary

17.20 Self-Assessment Test

17.21 Suggested Readings/Reference Material

17.22 Answers to Check Your Progress Questions

"Data! Data! Data! I cannot make bricks without clay."

- Sherlock Holmes, The Adventure of the Copper Beeches, 1892

17.1 Introduction

Information is a building block for business. Marketers should not only join the data revolution but should be leading the charge.

In the previous unit, we have studied the methods used in analytics at length. Here, in this unit, we will visit the application part of the analytics. We will

specially refer to business and marketing intelligence in this unit, as it is well associated with analytics. Business Intelligence (BI) can be used in different industries such as airline, retail, manufacturing, financial services, healthcare, bioinformatics and hospitality industry. The current day business intelligence systems are replacing DSS (Decision support systems), MIS (Management information systems) and EIS (Executive information systems). Organizations such as Tesco, Capital One, CEMEX and Netflix have made better decisions based on business intelligence.

In the current unit, the need for business intelligence and the definition of business intelligence are explained. The distinction between data, information, knowledge, and wisdom are explained. Data warehouse, business intelligence architecture, business intelligence components, business intelligence methodologies, data mining techniques, and business intelligence tools are described at length. The usage of business intelligence in market knowledge collection and its applicability to decision making is highlighted. Data in organizations is growing much faster than the computing speed in the world. Hence, the importance of big data, Hadoop and big data analytics are also explained in the unit.

17.2 Objectives

After going through this unit, you will be able to:

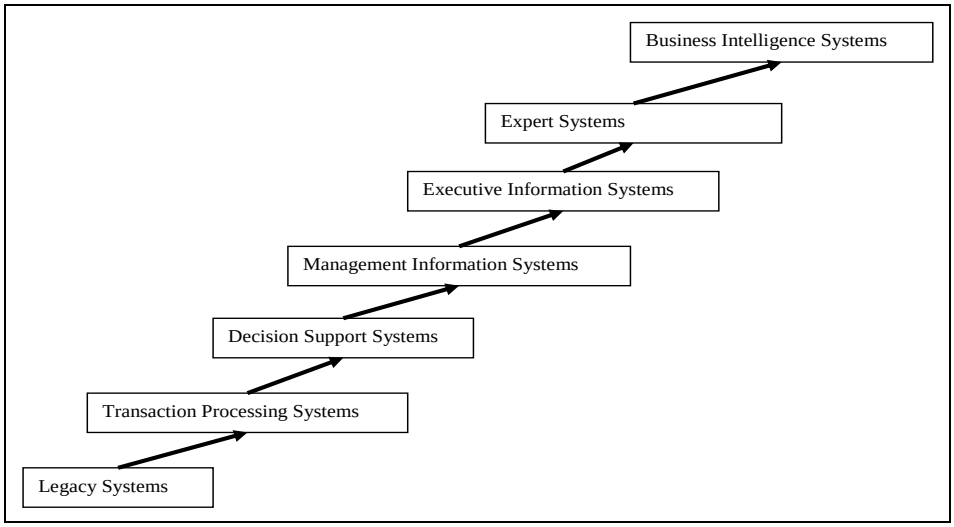
- Define different components of business intelligence
- Explain the business intelligence architecture
- Define data mining techniques used in business intelligence
- Explain the application of business intelligence in market intelligence and decision-making
- Discuss the utility of various business intelligence tools commercially available in the market
- Define big data architecture and Hadoop

17.3 Need for Business Intelligence

What is the need and application of business intelligence systems in an organization? In the absence of data for decision-making in an organization, it would be just guessing, instead of judging the current status of performance of a business. Hence, the organizations looking for performance improvements prefer to go for business intelligence. The organizations initially have to think why do they require business intelligence application? If they decide to have business intelligence, then they should think: which stakeholders get benefited using business intelligence tools and what are the investments for the acquisition of such applications? Business intelligence is required for effective decision-making, operational and strategic excellence.

Legacy systems used complex algorithms, being developed using procedural and functional programming languages and were mainly used in scientific computing. Transaction Processing Systems (TPS) used some form of data and file-based systems and these were used for business purposes. Decision Support Systems (DSS) used data models and interfaces for the users. Management Information Systems (MIS) used relational databases and business logic. Executive Information Systems (EIS) were used in management reporting and data visualization. Expert Systems (ES) were rule-based and search knowledge bases for analysis. The current day business intelligence systems (BI) use data mining techniques, data warehouses and business analytics tools useful for managerial decision-making. The evolution of information systems over a period of time is shown in Figure 17.1.

Figure 17.1: Evolution of Information Systems



Source: ICFAI Research Center

17.3.1 Overview of Business Intelligence

Business intelligence is the application of methodologies, processes and technology in acquiring, integrating, storing, accessing, analyzing and interpreting the data to make enterprise level decisions.

Business intelligence is about extracting the needed information and transforming that information into knowledge. Business intelligence extracts large amounts of data, analyzes it and generates reports needed for daily decision-making. Senior management and top management can be benefited from the insights and reports generated using business intelligence. Business intelligence technologies support efficient business operations. Business intelligence uses the technologies such as data warehouses, data mining tools, OLAP (Online Analytical Processing) tools, web services, XML, J2EE, and .Net. Business intelligence includes several software tools for extraction, transformation, load, querying, visualization and reporting.

Business intelligence is different from competitive intelligence. Competitive Intelligence (CI) concentrates only on the external factors of the organization, whereas business intelligence considers internal factors such as operational details of the organization as well. Business intelligence capabilities include data mining, online analytical processing, decision support system, forecasting, and statistical analysis.

Business intelligence facilitates effective communication in an organization. The organizations can change their strategies and decisions based on the changing economic conditions, customer preferences, product sales, financial situation, and supply chain operations using business intelligence. Using business intelligence, the organizations can find who their loyal customers, most profitable customers and potential customers are. One can also find out the reasons for customer loyalty using business intelligence. Business intelligence enables us to identify the business trends, anomalies, obtain insights and run simulations.

Example: Coca-Cola deployed "Business Intelligence" to Maximize Operational Efficiency

Coca-Cola is the largest independent bottling partner. The sales and operations staff did not have access to real time data to take decisions. The company had set up a BI platform for automating the reporting. The automation resulted in a saving of around 260 hours a year. The sales teams in the field could access mobile dashboards provided by the BI platform. The dashboards provided timely and actionable information. This gave a distinct competitive advantage. As the end users could generate their own reports and act on immediately, the operational efficiencies were enhanced tremendously. The IT team moved away from report generation to more strategic IT initiatives.

Source: https://www.tableau.com/learn/articles/business-intelligence-examples, 2022, Accessed on 01/10/2022

Activity 17.1

Business Intelligence: Retail Chain

Company X is a retail chain is operating in different locations in India. It collects data of products being sold, who purchased them, along with the customer profiles. The organization is currently collecting data based on customer surveys. But this process of finding results is taking a lot of time. The organization would like to know who their profitable customers are and what products are being mostly sold. The organization currently maintains a relational database. However, the top management needs a quick solution every time. What would you like to suggest to the organization? Which technologies, processes and approaches will solve their problem?

Answer:

17.4 Data, Information, Knowledge and Wisdom

The human brain contains four types of data. They are raw data, information, knowledge and wisdom. Data travels from information to knowledge and to wisdom.

Data: Raw data is the figures and numbers. Data alone cannot make any sense. It cannot give any meaning to the individuals. It has to be processed following certain rules to understand. The data in a spreadsheet or a flat file is an example of raw data. Data is frequently shared between the organization and other stakeholders of the organization such as customers, suppliers and partners. The characteristics of good quality data include completeness, correctness, timeliness and consistency.

Information is processed data. Status reports, trend reports and progress reports in the organization are processed data which gives information to the executives. The data tables with column names and row values in relational database management systems (RDBMS) provide useful information.

After reading the information, the individual understands it, interprets it and stores it. That becomes knowledge.

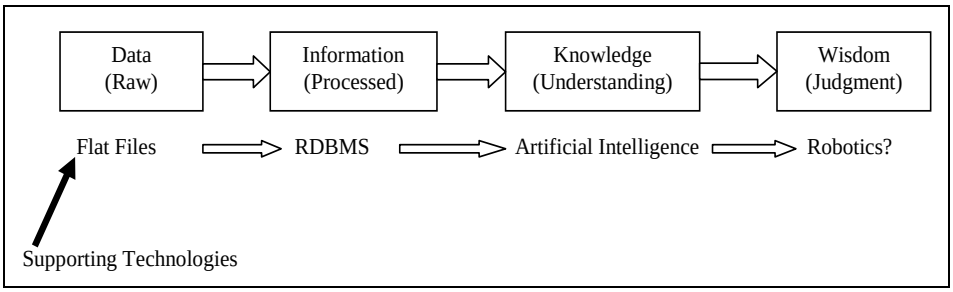
Knowledge is the information well understood. Further, the individual does not leave the understood information there itself. He applies his judgment, values and ethics into it and makes it wisdom. For example, remembering the number of defects per release of a software product, the number of test cases written per release of the product and making decisions based on that information is the knowledge of the product. This remembered knowledge is useful in the future journey of the product. Similarly, remembering the normal temperature of a human being and temperature of boiling water, etc., are examples of knowledge. Human beings remember this knowledge and apply whenever a need occurs. Artificial intelligence systems take decisions based on the compiled knowledge and logic. Expert systems search the knowledge bases.

Wisdom state represents the judgmental level of knowledge. Based on this, the individual takes decisions in the organization and decides whether it is right or wrong, acceptable or unacceptable, and ethical or unethical. Wisdom also involves future thinking and vision. As data transforms into information, knowledge and wisdom, the level of understanding increases in the individual.

Note: It is to be noted here that intelligence is an application of knowledge. Thus, a book contains information but the reader converts the concepts explained to intelligence and applying this knowledge contents to real life problems.

The journey of data from information to knowledge and to wisdom is shown in Figure 17.2 below:

Figure 17.2: Evolution of Knowledge



Source: ICFAI Research Center

Wisdom comes through systematic practice. Machines may not reach this level because judgment has to happen based on the facts, data, ethics, vision, values and culture. Inculcating or embedding ethics, culture and values into machines is still the grey area. Robots work based on the knowledge fed to it beforehand and artificial intelligence. Artificial Intelligence (AI) is the process of application of knowledge supplied to it. AI work as rule-based systems. AI systems do not have the wisdom as a human being has. Thus, Business Intelligence (BI) systems provide knowledge and information useful for decision-making in the organization. Business intelligence system is becoming part of knowledge management practices of the organization as well. They are useful not only for business decision-making but also for knowledge management in the organization.

Activity 17.2

Data, Information and Knowledge

A pharmaceutical company would like to improve the sales of its different products in rural areas. For this, their marketing strategy team recommended developing an expert system (ES) which can be deployed for this purpose. Using the proposed expert system, without a doctor's physical intervention, prescriptions can be generated by supplying the symptoms of the disease as input. The IT department is given the job of developing such an expert system. The project manager wonders whether to collect data, information or knowledge. Suggest how the project manager can go ahead with this project.

Answer:

Check Your Progress - 1

1. Business intelligence is required for which of the following?
   a. Organizational Performance
   b. Decision-Making
   c. Organizational performance and decision-making
   d. Loss of Productivity
   e. Knowledge
2. Business intelligence systems make use of which of the following?
   a. Data Warehouse
   b. Data Marts
   c. Data warehouse and data marts
   d. Data Loss
   e. Raw Data
3. What is information?
   a. Useful data
   b. Collected data
   c. Processed data
   d. Data about people
   e. Text without data
4. What does DSS stands for ?
   a. Demand Supply Support
   b. Divided Supply Systems
   c. Direct System Source
   d. Decision Support System
   e. Decision Supply System
5. Knowledge depends on which of the following?
   a. Understanding
   b. Withstanding
   c. Outstanding
   d. Application
   e. Data management

17.5 Data Warehousing

Data warehouse is the major component of business intelligence. It helps in the propagation of data in an organization. It extracts, cleanses, integrates, transforms and stores the data and transmits it for query processing and analysis as and when required. The sources of data for data warehouse can be the internal enterprise

systems, operational databases, relational databases, spreadsheets, historical databases, unstructured data and from point of sale terminals. The data from the internet and emails are also considered. It integrates the data required for organization's strategic, tactical and operational planning and decision-making. Data received can be in any of the following two forms: Structured and Unstructured Data (see Exhibit 17.1).

i) Structured data can be from the organization's relational databases such as tables, forms and spreadsheets. It is the data which can fit into an organizational relational database. Structured data is relatively easy to search.

ii) Unstructured data can be email messages, charts, graphs, memos, movies, images, telephone conversations, letters, news items, marketing flyers, presentations, spreadsheet files, web pages, whitepapers, discussion forum messages, pictures, biometrics (fingerprints, facial images), plain text files, audio and video files, etc. Some researchers have used the term semi-structured data to mean unstructured data. It is the data which cannot fit into a relational database or structured data. It cannot be represented in rows and columns. Semi-structured data analysis requires classification and taxonomy. It contains the important information needed for organizational decision-making. Data warehouse consolidates the data collected from various enterprise systems and the external data.

Exhibit 17.1: Teradata Solution for Financial Services

Techlogix is a company providing business intelligence solution to big companies worldwide. One of its clients, has locations across the globe. It has seven different data warehouses and business intelligence units working at different locations. It wanted to integrate all the data warehouses and have an enterprise data warehouse (EDW) which could cater to the needs of its financial services business. Techlogix took up the job and implemented the solution for the client, using Teradata, a data warehousing platform. The enterprise data warehouse was implemented using technologies such as Teradata, Teradata ETL tools and Informatica. Business objects were used for front-end development and report generation purposes. Data from four different legacy systems and three different reporting environments were consolidated and extracted into the enterprise data warehouse using ETL tools. The Techlogix team had done data modeling, data extraction using ETL code, reporting, and front-end generation using Business objects. The data was modeled based on different subject areas such as events, party and agreement. Data modeling standards such as naming and typing were followed. They had developed ETL code for 10 different data sources. ETL code was developed using Teradata BTEQ and Informatica.

Contd…

The data sources for ETL tools used were Siebel, PMSVFS, ACBS, Workflow, Strategy, REM, GEODE, and RDM. They had followed the ETL coding standards. These ETL scripts extracted data and stored in an enterprise data warehouse. They had generated 100 different reports. They used analytics in the areas such as enterprise marketing, financial reporting, CRM and customer prospecting. They developed a dashboard and enterprise sales compass application.

Source: Techlogix (Undated), "Business Intelligence for Financial Services: A Case Study," Available online at http://www.techlogix.com/PDFs/BI%20for%20Financial%20Services.pdf, pp. 1-5.

Example: Uber uses PrestoDB (Open-source Database) to Enhance Flexibility in "Data Warehousing" and Querying

Uber has recognized that increasing size and complexity of its enterprise data needs a technology that can handle the various data storage units and various data formats. The company chose the open-source query engine PrestoDB. This choice will enhance the company to see large scale improvements in large-scale business intelligence and analytics,

It also allows Uber to have a single-entry point for its end users to run their queries and get insights rather quickly. Presto also works easily with the other open-source data and storage formats like Hadoop, Hive and Spark.

Source: Uber uses Presto to improve flexibility in data warehousing and querying - SiliconANGLE, March 29, 2022, Accessed on 07/10/2022

17.6 Business Intelligence Components

The essential components of business intelligence systems include data warehouse, data marts, corporate performance management systems, ETL tools, OLAP, analytical tools, data visualization, data mining, geographic information system and a well-defined workflow. Data warehouse is the important component of business intelligence. However, it should be a real-time data warehouse.

Data mart is an organized collection of data specific to given departments. It is a subset of a data warehouse. That is, a data mart is formed by extracting data from a data warehouse based on the department, specific business function, business process or business unit. This is helpful in making decisions specific to that department. For example, there can be different data marts for marketing, sales, finance, operations, and HR. Each data mart is useful for efficient decision-making for that department. There can be multiple data marts in one enterprise. Each data mart is formed to achieve operational excellence through decision-making. Functional executives can take decisions based on data

extracted from the data mart. Virtual data marts can also be created using database 'views'. Cubes are to be created from a data mart.

Corporate Performance Measurement can be done using organizational web portals, dashboards and scorecards. Key performance indicators (KPI) are also components of business intelligence. KPIs are the metrics collected weekly, monthly, quarterly, and yearly in the organization. Extract, transform, and lad (ETL) tools are also components of business intelligence. Current day ETL tools extract the data very quickly.

Activity 17.3

Business Intelligence Components

A manufacturing company has many locations with location-specific databases. The organization would like to consolidate all the data and would like to have consolidated reports. For that purpose, the company decided to go for a centralized data warehouse with business intelligence capabilities. The business intelligence objective is to acquire data, organize data and analyze data. In that direction, the project manager thinks of how to acquire data from different sources. Suggest to the project manager how he can extract data from different data sources. Suggest certain commercially available tools for this purpose.

Answer:

17.6.1 OLAP

OLAP (online analytical processing or OLAP server) provides multi-dimensional views, analyzing, visualizing, reporting and modeling the data. They can be used to optimize business operations. They work with data warehouses and data marts in course of accessing the data. They process queries which are needed to find the trends in the organization. Current day OLAP tools access the data and generate reports very quickly. OLAP tools take 0.1% of the time that a traditional relational database system takes for answering a query. Popular OLAP tool vendors include Cognos and Business Objects.

Analytics tools do the statistical analysis needed for forecasting, data mining and predictive analysis. They predict or provide insights based on certain facts for the

organization. The business intelligence components include business process model, business function model, business data model, metadata repository and application inventory.

Example: Netflix uses the OLAP Querying Functionality to Gain Insight Into how the Network is Behaving and Performing

Netflix is the world's leading internet television network, with more than 100 million members worldwide enjoying 125 million hours of TV shows and movies each day.

Netflix Uses real-time logs from playback devices to get insights into, to understand and quantify users' devices which are handling browsing and playback. It is very inefficient to load such huge volume of data (2 million events per second) into traditional data bases to query and get insights. Netflix team instead uses Druid (a high-performance analytics data base). Once into Druid, the data is analysed using OLAP tools.

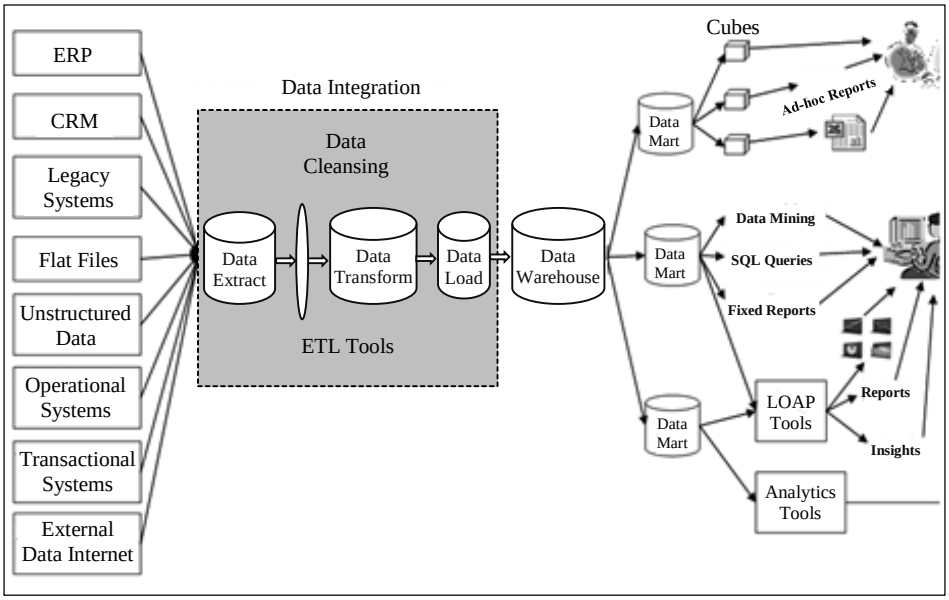
Source: https://netflixtechblog.com/how-netflix-uses-druid-for-real-time-insights-to-ensure-a-high-quality-experience-19e1e8568d06 date 03/03/2020, Accessed on 01/10/2022

17.7 Business Intelligence Architecture

The objective of business intelligence (BI) systems is to provide quality and timely input data for decision-making in an organization. They provide information on demand. They combine the operational systems and data with analytical tools in order to provide complete information for decision-making and planning in the organization. Business intelligence architecture consists of the data sources, data integration, data storage, data management, operational processes, presentation tools and applications, querying, and reporting as shown in Figure 17.3. This architecture should fit into the enterprise system's architecture. Business intelligence architecture is also part of the enterprise system's architecture. It is not an isolated entity in the organization. It is but a part of the enterprise network of systems.

The data sources for business intelligence are heterogeneous. It includes both internal and external data sources. Internal sources of data for business intelligence include enterprise resource planning (ERP), customer relationship management (CRM), legacy systems, flat files, operational systems, decision support systems, executive information systems, knowledge management systems, OLAP, visualization systems, transaction systems, and geographical information systems. The unstructured data used are messages, video, audio and external sources of data include the Internet, e-mails, blogs, social networking sites, and media. Data warehouse pulls the data from all these sources. The data type can be structured or unstructured data.

Figure 17.3: Business Intelligence Architecture



Source: ICFAI Research Center

Data integration is done through data extraction from different sources, data cleansing, data transformation into a required format, and then data loading into the data warehouse. Here ETL tools known as extract, transform and load tools can be used for data integration purpose. There are many commercially available ETL tools in the market. Some of the ETL tool vendors include Informatica, Trillium, Ascential, and Ab Initio.

Data warehouse provides the access, storage and integration to the data. The data from the data warehouse is loaded into data marts specific to the business function. Data mart is the tiny database specific to a department, business unit, business process, or business function. The advantage of data mart is that it provides quick access to data for specific purposes of the group. There can be multiple data marts in the organization. Data marts can also be used for SQL querying, fixed report generation and data mining purposes. There are different data mining techniques which can be applied on data marts. The outputs are pulled from the data marts.

Cubes are derived from data marts. A cube can be a logical view of the data. It provides structured information to the users. It is useful for querying and reporting purposes. The developers can derive multiple cubes from a single data mart. The developers and users access the data cubes. Data cubes can be used to generate ad-hoc reports.

OLAP (online analytical processing) tools and other analytical tools can access the data marts in the business intelligence framework. These tools apply statistical techniques and derive insights and findings useful for managers. They can also

be used to generate graphs and reports from the data. They can even generate trend reports, progress reports and status reports.

Overall business intelligence makes use of both internal and external data, analyzes it and prepares reports, graphs, insights and knowledge useful for decision-making at different levels in an organization. Business intelligence systems should be transparent, reliable, accessible and secure. They should be able to handle different data types, data formats and data sources. Business dominates the technology in business intelligence architecture. Business intelligence architecture also includes the metadata, standards, business rules and policies. Technical architecture consists of hardware, database management systems and middleware. The security and scalability of the business intelligence systems are also to be taken into consideration while architecting business intelligence for the organization. Business intelligence architectures should comply with the regulatory requirement of the Sarbanes-Oxley Act of 2002.

Metadata repository contains details about the source of data, bibliographic information, data definition and how it is processed. It also contains details about the reliability and accuracy of data.

Example: Intercom Adopts Data Integration Tools to Reduce Data Cleaning Effort from 10 Hours a Week to One Hour

Intercom, a California based Business Messaging Software Company, specializes in business messaging, providing businesses with a way to chat with their customers. The company had data in different places, and it used to take around 10 hours per week to extract data for analytics. The company deployed data integration tools to reduce this to one hour per week. The tool provided transparency for data while data was getting extracted . This enhanced the company's trust on data. The company is able to integrate product, marketing, sales, and pre-sales data with financial data to support decision-making.

Source: https://www.fivetran.com/case-studies/case-study-intercom company case study, Accessed on 01/10/2022

Check Your Progress - 2

1. Which of the following can be a data type?.
    a. Structured data
    b. Unstructured data
    c. Structured or unstructured data
    d. Error data
    e. Scattered data

2. Data warehouse functionality includes which of the following?
   a. Cleansing data
   b. Storing data
   c. Transforming data
   d. Restructuring data
   e. Extracting data
3. Business intelligence architecture includes which of the following?
   a. Data warehouse
   b. ETL tools
   c. Data marts
   d. Data warehouse, Data marts, ETL tools
   e. Software
4. What should be business intelligence architecture?
   a. Scalable
   b. Secure
   c. Scalable, secure
   d. Unreliable
   e. Loosely coupled
5. ETL stands for which of the following?
   a. Extract, Transform and Load
   b. Enter, Transfer and Leave
   c. Early Transfer Level
   d. Electronic Transfer and Leave
   e. Execute, Translate and Leave

17.8 Business Intelligence Methodologies

Business intelligence can be viewed as an application of data mining technique, usage of complex algorithms and statistical analysis on data. Business intelligence methodologies include predictive analysis, statistical analysis, reporting, and ad-hoc analysis. Business intelligence involves the detailed analysis of huge data, application of technologies and analysis practices.

The methodology to deal with structured and semi-structured data in business intelligence is to acquire the data, clean up the data and integrate the data. Then, search the data, analyze the data, identify trends, changes and incorrectness, and deliver the results. The management action is based on the provided information and results. The reports generated can be validated, structured and summarized. For structured data, analysts use ETL tools, data warehouse, OLAP, and data

mining. Semi-structured data requires different and less sophisticated tools to analyze. Semi-structured data can be gathered from business processes and news items.

Complex Analysis makes use of fast and user-friendly OLAP queries. OLAP queries are used in marketing, financial reporting, business process management, sales and forecasting. The OLAP analyst traverses through the data warehouse, data marts and changes the data orientation. The operations possible in OLAP include slice and dice—a capability to combine and recombine different combinations of data; drill down/up —to navigate through data, pivot which changes the dimensions from rows to columns and vice-versa; nesting —displays one dimension inside another dimension of data.

Business intelligence analytical techniques include modeling, visualization, embedding, monitoring, reporting, data mining, scorecards and investigating.

Example: HelloFresh (Meal Kit Company) uses Centralized Digital Marketing Reporting to Increase Conversions

HelloFresh supplies fresh food and recipes directly to consumer households in ten international markets. Everything needed for the weekly meals is meticulously organized for each subscriber, sourced locally, and delivered to the doorstep at the desired time. It was based on a subscription model with any-time cancellation.

HelloFresh deployed the Tableau platform to centralize global performance reporting, saving 10–20-man hours per day and providing regional sales and marketing teams with real-time data for nimble decision making. As a result, the HelloFresh marketing team can react to trends in customer behaviours and optimize marketing campaigns on-the-fly, leading to better conversion rates and improved customer loyalty.

Source: https://www.tableau.com/learn/articles/business-intelligence-examples company case study, Accessed on 01/10/2022

17.9 Data Mining Techniques

Data mining is the process of finding patterns in huge data using statistical techniques, artificial intelligence and Database management system (DBMS). Data mining can be used in fraud detection, marketing and supervision. It finds correlations, patterns and trends in data from data warehouse using statistical and mathematical techniques. Data mining can be used for hypotheses proving and knowledge creation.

The main objective of data mining is to find the earlier undetected patterns in large data sets (of the organization). Data mining techniques include classification, multidimensional analysis, correlation, regression, associations, prediction, clustering, time series analysis and outlier analysis. Exploratory data analysis and sequential pattern analysis are other data mining techniques.

Classification determines the characteristics of a particular group. Each group characteristics can be used to design a model. Clustering creates the group of observations having certain common characteristics. Time series analysis finds the associations based on time. Association finds the relationship between events. Correlation finds the relationship between two different variables or events. Regression finds the impact of one event on other event. Regression is of two types such as linear regression and non-linear regression. Prediction finds the future values based on huge data sets.

Multidimensional analysis requires a multidimensional database. Multidimensional analysis can be done on three-dimensional cubes. A Cube can also be called as a multidimensional database. Cubes are useful to generate ad-hoc reports and ad-hoc queries. The multidimensional setup with cubes should support pivot analysis useful for generating ad-hoc reports. An example of a three-dimensional cube consists of Products, customers and time of product purchased. Master data is used to generate different dimensions in the database. Master data is important for business intelligence since it can be drilled to facilitate analysis and creation of various other data and information. Examples of master data include customer master file, product master file, supplier master file, etc.

Fixed reports can also be generated using multidimensional cubes. However, this process does not use standard SQL (structured query language), but it uses a multidimensional language known as Multi-Dimensional Expressions (MDE). Hence, the best practice to generate fixed reports is to use SQL commands directly from data warehouse and use cubes for multidimensional reports. Data mining tools are based on artificial intelligence, statistical and mathematical techniques, decision trees, neural networks, and Bayesian network theory. Commercially available data mining tools include IBM Intelligent Miner, SAS Enterprise Miner, DBMiner, R, SGI Mine Set, and MS SQL Server. There are some text mining tools as well.

Example: JP Morgan Chase uses "Data mining "to Speed Up Loan Application Processing With Reduced Errors"

JP Morgan Chase has introduced Contract Intelligence (COIN) to process customer applications. The COIN system is based on Data mining and Natural language processing. The bank processed around 12000 applications a year. Usage of COIN has reduced the effort by 360 thousand hours. Also, errors were reduced. The detailed and deep analysis of all aspects of the applicant obtained from different sources means the company has a very good insight into the credit risk of the applicant.

Source: https://algorithmxlab.com/blog/data-science-8-powerful-applications/ date August 28, 2021, Accessed on 02/10/2022

Check Your Progress - 3

6. OLAP stands for which of the following?
    a. Online Application Processing
    b. Online Application Performing
    c. Online Analytical Processing
    d. Offline Analytical Procedure
    e. On Load Analytical Procedure
7. Data mart is _____.
    a. Sub-set of data warehouse
    b. Specific to business function
    c. Specific to department
    d. Specific to department, business function and subset of data warehouse
    e. Total database
8. Cubes are derived from which of the following?
    a. Data Fields
    b. Data Marts
    c. Data Entries
    d. Data Summaries
    e. Storage Blocks
9. Business intelligence methodologies include which of the following?
    a. Predictive Analysis
    b. Statistical Analysis
    c. Ad-hoc Analysis
    d. Predictive, statistical and Ad-hoc analysis
    e. Risk Analysis
10. Which of the following is not a data mining technique?
    a. Classification
    b. Multidimensional analysis
    c. Clustering
    d. Cloning
    e. Factoring

17.10 Market Intelligence and Decision Making

Business intelligence implementation in an organization involves hardware, software, human resources, and costs of implementation. Training is also required

for analysts to use the system. System upgrades also initiate training need in the organization. Business intelligence can be used in deriving competitive intelligence for the organization. Competitive intelligence is about gathering and analyzing external information useful for devising organizational plans, strategies, operations, and decisions. Market intelligence includes competitive intelligence, competitive strategies, pricing strategies, sales strategies and competitive advantages.

The sources of market intelligence include government websites and portals, online databases, government publications and reports, online databases, surveys, trade associations' periodicals and reports, user groups, consumer groups, industry bodies, industry consortiums, competitors, suppliers, vendors, partners, customers, distributors, interviews with industry experts, journals, newspapers, magazines, financial reports and private sector organizations.

The data collected from the above mentioned sources is to be fed into the organizational business intelligence system to gain market intelligence in the industry. The output reports, graphs, knowledge and information of business intelligence are useful in organizational decision-making.

Example: Univision Increases Market Spend Efficiency through "Marketing Intelligence" Based Decision Making

Univision is an American Spanish-language television network. It's the largest provider of Spanish-language content in USA.

Univision faced the challenge of not having visibility on spend targeted on a campaign.

The company introduced "Programmatic TV" .This is a data based automated method for deciding on the ad campaigns and spends. Using BI tools, data from applications like Facebook, Google Analytics and Adobe Analytics are analysed and the company gets valuable insights for the right and effective ad spends.

The company realized an 80% growth in yield during the first quarter after implementing business intelligence.

Source: https://www.netsuite.com/portal/resource/articles/business-strategy/business-intelligence-examples.shtml Company case study, Accessed on 02/10/2022

17.11 Make the Last Mile in Data Analytics

The 'last mile' is the group of people who deliver the results of the data analysis. The group gives this result to the business so that they can easily understand the trend. The 'last mile' group has expertise in data analytics and knows enough about the business. It requires experience in data analytics and also the confidence to present the results to the CEO. The 'last mile' group can handle big issues and help in developing and guiding business strategies.

For example, Yahoo Mail has 250+ million users. When people sign up for Yahoo Mail account, they see the news preview module first. The news preview module has become popular because it helps in retaining active users. Data analytics noticed that new users like to read the news when they read email. By adding a news preview window, Yahoo mail was able to increase the return rate by 40%. In addition to new users, the other users also liked reading news while looking at their email.

17.11.1 Geospatial Intelligence

Geospatial intelligence means using data related to space and time to improve the quality of predictive analysis.

For example, the smartphone helps to look at traffic and it shows streets in red and yellow color. It observes the average speed of travel and calculates the aggregate speed to travel and then it helps us in avoiding traffic. Therefore, geospatial analytics has become a standard part of life today.

For advertisers, geographical intelligence helps in a different way. It makes the users feel ads less like spam and more like information. The following are the examples of geospatial intelligence:

- Healthcare organizations will be able to predict movements of disease outbreaks over time and adequately prepare for potential epidemics before they occur.
- Police departments can study past geospatial data to see where crimes occurred frequently and understand where and when future crimes are most likely to happen.
- Insurers can incorporate geospatial information into their risk calculations to optimize pricing for known risk factor.

17.11.2 Consumption of Analytics

Consumption of analytics means making analytics consumable in an organization. There are different stages in the consumption of analytics. They are:

i) Communication

ii) Implement

iii) Measure

iv) Align incentives

v) Develop cognitive repairs

i) Communication

In the first stage, the business analytics from the core team will be sent to the wider group of decision makers and the daily consumers of analytics in your organization. It helps the team to create a platform for analytics in an organization.

ii) Implementation

Implementation means to get all the ingredients in place to consume analytics successfully. Strong leadership can be the most important trigger in adapting analytics in an organization.

iii) Measurement

Measurement means testing of consumption. It uses analytics to test itself. A successful business decision can be taken only with a combination of business experience and analytics.

iv) Align incentives

Successful consumption of analytics leads to the creation of structured decision-making processes which is produced by data analysis.

v) Develop cognitive repairs

Creation of business insights based on data and then going and proving it right for all to see is by far the most effective to both expose biases and create repairs.

17.11.3 From Creation to Consumption

Various organizations have created analytics but have failed in consumption. Creating analytics does not automatically result in consumption.

The following are some key questions:

- Do you have experience in creating analytics but failed in consumption?
- Do you have any problem in maintaining the balance between analytics creation and consumption?

If the answer to any of these questions is 'yes', that means your organization suffers from the creation-consumption gap. Organizations should be able to manage this creation-consumption gap and capitalize analytics as a source of consumptive advantage.

17.11.4 Analytics for Business and Market Intelligence

Big data analytics uses three types of business analytics. They are:

i) Descriptive analytics

ii) Predictive analytics

iii) Prescriptive analytics

i) *Descriptive analytics:* It describes the previous business analytics. It uses SAS and SPSS for descriptive statistics.

ii) *Predictive analytics:* It uses the previous business analytical information and predicts future outcomes with some degree of likelihood.

iii) *Prescriptive analytics:* It uses previous business information to direct future activities to achieve optimal results.

These three techniques have been used for decades, combining with big data in shifts. Some of the important factors/aspects to be considered while dealing with analytics include:

- Using more or all of the data for predictive model.
- Combining analytical models to improve the results.
- Using new learning in predictive models.
- Making predictive model close to real-time analytics.
- Applying predictive models rather than new techniques.

Table 17.1 discusses three different types of analytics models.

Table 17.1: Diagnostic vs. Descriptive vs. Predictive vs. Prescriptive Analytics

The four types of advanced analytics defined according to their differences. Below is a summary of their operations:

Diagnostic	Descriptive	Predictive	Prescriptive
Uses historical data	Uses historical data	Uses historical data	Uses historical data
Identifies data anomalies	Reconfigures data into easy-to-read formats	Fills in gaps in available data	Estimate outcomes based on variables
Highlights data trends	Describes the state of your business operations	Creates data models	Offers suggestions about outcomes
Investigates under-lying issues	Learns from the past	Forecasts potential future outcomes	Uses algorithms, AI and machine learning
Answers "Why" Questions	Answer "What" Questions	Answers "What Might Happen?"	Answers "If, Then" Questions

Source: Allison, M. (2021, May 12). What Are The 3 Types Of Business Analytics. Darrin Kenney's Templates. https://jerseystrife.blogspot.com/2021/04/what-are-3-types-of-business-analytics.html

Example: Intangles Enables One of its Customers to Achieve 85% Increase in Safety Using Digital Technologies Including "Geo Spatial Intelligence"

Intangles, a digital twin technology start-up company, is a start-up company based out of Pune with an office in USA. The company is in the business of intelligent mobility solutions based on Digital Twin technology and predictive analytics. One of its clients is a pan India logistics company with a network of commercial vehicles. The company faced a challenge in not being able to control safety incidents involving drivers and vehicles. \

Contd….

The management was looking for a solution to enhance overall safety score of the drivers and vehicles there by reducing accidents. The management also desired real time dashboards on driver and vehicle performance to take real time actions.

Intangles suggested and implemented a solution which is based on Digital Twin Technology and using analytics algorithms including "geospatial intelligence". The solution provided greater visibility on various aspects of its fleet management operations. It also developed an 'Inline Driving Scorecard' feature. Through these insights, the company improved vehicle safety and overall efficiency in a short period of time. Top performing drivers got incentives, and poor performers were trained.

Source: https://www.intangles.ai/case-studies/85-percent-jump-in-vehicle-safety-through-data-driven-insights/ company case study Accessed on 02/10/2022

17.12 Correlation Analysis

What is correlation?

Correlation is the most useful statistics which describes the degree of relative-predictive-and-prescriptive-analytics.

After calculating correlation, we determine the probability of observed correlation by conducting a test of significance. If one thing causes another, then we say that the two will be correlated. Two things that are correlated are not necessarily related by cause because one is a subset of another.

Big data brings in great responsibility. Advanced algorithms help us to analyze vast amounts of data. We must continue to rely on the expertise of data scientists to ask the right questions and draw the correct conclusions.

17.12.1 Logistic Regression

Logistic regression is a statistical method for analyzing a data in which there are one or more independent variables that determine an outcome are.

- Logistic Regression is a predictive model.
- Logistic regression model does not involve decision trees.
- Logistic regression can be used only with two types of target variables:
- A categorical target variable
- A continuous target variable

17.12.2 Factor Analysis

Factor analysis is a tool used to measure the relationship between large numbers of variables. It allows researchers to use psychological scales to measure directly by collapsing a large number of variables.

The main concept of factor analysis is to measure the variables which are associated with a latent variable (which is not measured directly). For example,

people may respond similarly with regard to income, education and occupation—all of which are associated with latent variable socioeconomic status.

In every factor analysis, the number of factors and variables are the same and the factors are always listed in the order of variation. Therefore, each factor captures overall variance in the observed variables.

The eigen value is a measure of the variance of the observed variables. Any factor with an eigen value ≥1 explains more variance than a single observed variable.

Factor loading

Factor loading means the relationship of each variable under each factor. Here is an example of the output of a simple factor analysis with just six variables and two resulting factors. Table 17.2 shows example factor loadings of different factors.

Table 17.2: Factor Loadings

Variables	Factor 1	Factor 2
Income	0.65	0.11
Education	0.59	0.25
Occupation	0.48	0.19
House value	0.38	0.60
Number of public parks in neighborhood	0.13	0.57
Number of violent crimes per year in neighborhood	0.23	0.55

Source: Rahn Maike (2013). Factor Analysis: A Short Introduction, Part 1. The Analysis Factor (2013). http://www.theanalysisfactor.com/factor-analysis-1-introduction/

The variable with the strongest association with the underlying latent variable Factor 1, is income, with a factor loading of 0.65. So that we can say that the variable income has a correlation of 0.65 with Factor 1. This would be considered a strong association for a factor analysis in most research fields.

17.13 Marketing Intelligence Using Analytics

Today, leading companies are looking to improve business performance via faster, better decision-making by applying advanced predictive modeling to their vast and growing volumes of data. Business analytics, whether for marketing, CRM, loyalty or operations, provides organizations with valuable insights from their data — allowing them to uncover and act on new opportunities to increase revenue and profitability.

Market basket analysis is a data modeling technique used to find associations between items or events by determining the likelihood for them to occur together. Taking its name from the concept of identifying products that customers are putting into their shopping cart, Market basket analysis is also commonly referred to as product affinity analysis or association rule learning.

Market basket analysis is one of the many advanced models. A typical approach of market basket analysis is to leverage large amounts of customer transaction

data to determine products that are most commonly purchased together. Understanding these purchasing patterns empowers marketing and sales organizations to make more informed decisions about how and where to deploy their efforts and resources.

An obvious application of market basket analysis is in the retail sector where retailers have large amounts of transactional data and often thousands of products. One of the recognizable examples is the Amazon.com recommendation system: "Customers who bought this item also bought items A, B and C".

In order to capitalize on big data value, big data apps have started to emerge. The horizontal big data apps (machine log analytics) and vertical big data apps (telecommunications analytics) are emerging.

These emerging techniques are designed to solve specific business problems which incorporate deeper and more complex prescriptive analytics.

Top emerging technologies include:

- Fuel-cell vehicles (Cars that run on hydro zed)
- Next generation robotics (Rolling away from the production line)
- Recyclable thermoset plastics (A new kind of plastic to cut landfill waste)
- Precise genetic-engineering techniques (A breakthrough; offers better crops with less controversy)
- Additive manufacturing (Making things from printable organs to intelligent clothes)
- Emergent artificial intelligence (What happens when a computer can learn on-the-job?)
- Distributed manufacturing (The factory of the future is online and on your doorstep)
- Neuromorphic technology (Computer chips that mimic the human brain)

Example: Verizon uses "Telecommunications Analytics" to Identify New Revenue Streams

In the highly competitive telecommunications industry, remaining relevant to customers and identifying new sources of revenue is critical, especially when the current revenue sources are in decline. Data analytics and Business Intelligence have facilitated the company becoming the nation's largest and most reliable network provider. Telecommunications analytics is also facilitating the company's future success. R&D teams used data, analytics, and strategic partnerships develop offerings, related to the Internet of Things (IoT). The new dimension of data is IoT. It will open up to new revenue streams. Smart cars, smart agriculture, and smart IoT will all be part of this new growth.

*Source : https://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB9520.pdf?_gl=1*p98ipi*_ga*NTc5MjE3NTIyLjE2NjQ3MTA2MDc.*_ga_7PE2TMW3FE*MTY2NDcxMDYwNy4xLjEuMTY2NDcxMDYxNy4wLjAuMA company case study, Accessed on 02/10/2022*

17.14 Customer Experience Management Using Analytics

Market basket analysis (MBA) is a data mining technique which looks at purchase coincidence and is widely used in the consumer packaged goods. It investigates whether two products are being purchased together and whether the purchase of one product increases the likelihood of purchasing the other.

Market basket analysis results in a better understanding of your customers and their purchasing behavior by allowing you to explore associations and predict the likelihood of a customer's future purchase behavior based on associations. It is one of many advanced business analytics tools that can help organizations optimize marketing and sales operations for improved performance.

Marketing and sales organizations across all industries are looking to analyze, understand and predict customer purchase patterns towards achieving strategic goals to reduce churn rates and maximize Customer Lifetime Value (CLV). Selling additional products and services to existing customers over their lifetime is the key to optimizing revenues and profitability. Market basket analysis association rules identify the products and services that customers typically purchase together, empowering organizations to offer and promote the right products to the right customers.

Moreover, with predictive analytics, organizations are able to promote their most profitable products and services to the most likely buyers. They can also encourage additional purchases by introducing new targeted products, products with high margin, or high performing products which may not have otherwise been an obvious next purchase.

Example: Eureka Forbes Engages with Customers Throughout the Year to Ensure Customer Loyalty and Hence Higher "Customer Lifetime Value"

Eureka Forbes engages with its customers round the year that will ensure high "customer lifetime value". The company gets individual customer view based on data from across channels. The data is analysed, and the insights are passed onto everyone in the chain so that the appropriate response can be given to the customer. The company understands the product it is marketing is robust so the customer experience should be outstanding in the first place. The company maintains low customer complaint rate, provides post purchase demonstration, and thereby enriches customer experience. The customer is retained for life.

Source : https://www.businessworld.in/article/Present-Day-Contours-Of-The-D2C-Model/02-09-2021-402919/ date o2/08/2021, Accessed on 02/10/2022

17.15 Business Intelligence Tools

Business intelligence tools include AQL (associate query logic), decision support systems (DSS), executive information systems (EIS), management information systems (MIS), query and reporting tools, OLAP (online application processing) tools, data mining tools, and ETL (extract, transform and load) tools.

Most influential commercially available business intelligence tools are from organizations such as Business Objects, Microsoft, SAS, Teradata, PeopleSoft, ORACLE, IBM, Manhattan Associates, Insight Software and OutlookSoft.

Statistic on Growth of BI

- Over 50% of business users and analysts will have access to self-service tools in 2017.
- 42% of companies plan to utilize mobile business intelligence.
- Companies using analytics are five (5) times more likely to make faster decisions.
- Customer analytics (48%), operational analytics (21%), and fraud & compliance (21%) are the top three use cases for big data.
- There will be a projected 1.5 million shortage of data professionals in the United States alone by the year 2018.
- 95% of large organizations will hire a Chief Data Officer by 2019. (source)
- 85% of business leaders believe big data will change the way they do business. (source)
- 89% of business leaders believe big data will revolutionize business operations in the same way the Internet did. (source)
- By 2019, the business intelligence and analytic market will grow to $20 billion. (source)
- Wikibon projects the Big Data market will top $84B in 2026, attaining a 17% Compound Annual Growth Rate (CAGR) for the forecast period 2011 to 2026. (source)

Source: Ofori-Boateng Christian (2020.) Everything You Need to Know About Business Intelligence. Christians even BI Blog (Jun 10, 2020)

https://go.christiansteven.com/bi-blog/informative-stats-the-growth-and-value-of-business-intelligence

Oracle business intelligence applications include Oracle Financial Analytics, Oracle Project Analytics, Oracle Sales Analytics, Oracle Price Analytics, Oracle Marketing Analytics, Oracle Procurement and Spend Analytics, Oracle Supply Chain and Order Management Analytics, Oracle Human Resources Analytics, Oracle Service Analytics, Oracle Loyalty Analytics, and Oracle Call Center Telephony Analytics. Oracle business intelligence applications are capable of integrating with Oracle E-Business Suite, JD Edwards Enterprise One, PeopleSoft Enterprise and Siebel CRM. It consists of more than 3,000 pre-built reports. A proper analytics package comes with data schemas, dashboards, predefined reports, business views, and an integrated set of tools. The business intelligence tools for semi-structured data are still maturing.

17.15.1 Big Data

The amount of data growing in organizations is huge and also unmanageable sometimes.

Big data can be used in wide areas such as retail, mobile services, e-commerce, education, financial services, scientific research, manufacturing, life sciences, bioinformatics, physical sciences and astronomy. Big data applications include traffic management, urban planning, environmental modeling, smart materials, computational social sciences, financial risk analysis, security and intelligent transportation, seismic data analysis, meteorological data analysis, etc.

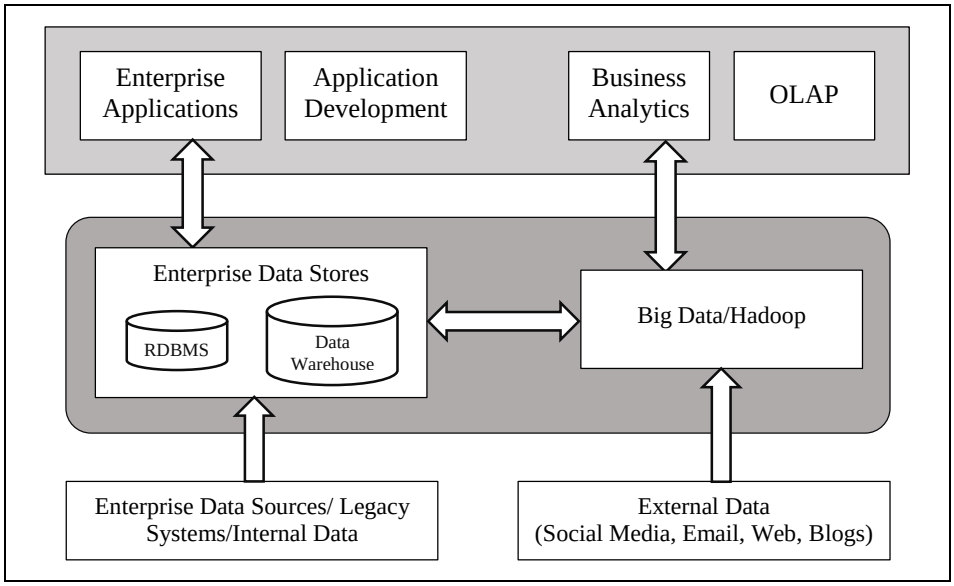
Do You Know?

In 2002 itself, $31billion e-mail messages were sent across the world. In 2006, it was 60 billion e-mail messages and in 2015 it was 205 billion.

Source: Craig Smith (2020).90 Interesting Facts About E mail/How Many Emails are Sent Per Day? DMR Business Statistics Fun Gadgets (July 2020). http://expandedramblings.com/index.php/email-statistics/

Organizations may use big data for customer retention, risk assessment and brand management. Big data requires different data mining techniques other than the traditional statistical techniques. Big data analysis process is explained in Figure 17.4.

Figure 17.4: Big Data Architecture



Source: ICFAI Research Center

17.15.2 Hadoop

Apache Hadoop is an open source platform to manage big data. It addresses the issues such as low cost, reliable storage and tools for analyzing unstructured data, etc. It is a project of Apache Software Foundation. Hadoop consists of a fault-

tolerant system known as Hadoop distributed file system (HDFS). It provides storage infrastructure that can hold data without loss. It creates clusters of machines and coordinates between them. These clusters are built with less expensive hardware machines. HDFS stores three copies of each block of data in three different servers in the cluster. Even if two servers go down in the cluster, the file can still be retrieved without any data loss.

Hadoop tool works at the whole quantity of data in the relational database. It uses a technique known as MapReduce that splits the task and pushes it on to different servers and later collates the results achieving the operational parallelism. It automatically restarts the work if any node goes down in the cluster. Hadoop distributed file system (HDFS) and MapReduce are the key features of Hadoop providing a reliable and low-cost solution to big data

17.15.3 Big Data Analytics

An organizational business analytics can include customer analytics, supply chain analytics, IT analytics, HR analytics, financial analytics, etc. For example, customer analytics can find what type of customers are profitable to the organization, supply chain analytics can find what inventory optimization levels are suitable for the organization, IT analytics can find whether IT services are efficient in the organization or not, HR analytics can find out what is the cost of recruitment, financial analytics can find what are the cost and revenue drivers of the organization and their impact on the profitability.

Do You Know?

In an IDG (International Data Group) survey, it was found that the Return on Investment (ROI) of business analytics in organizations ranges from 17% to 2,000%. The average return on investment of business analytics in organizations is 457% (2013).

Source: https://www.idg.com/news/idc-study-finds-analytics-projects-yield-431-average-roi/

Some of the analytical technologies include neural networks, genetic algorithms, swarm intelligence, information extraction, text categorization, text mining, audio mining, video mining, rule-based engines, data mining tools, simulation tools, spreadsheets, and OLAP tools.

Example: Bank of America unlocked $50M in Potential Revenue by Using Fiddler to Align "Machine Learning" Models With Business Context

Bank of America's decision-making process is based on descriptive and prescriptive analytics derived from Machine Learning models. But the value of insights from ML models decreases when the models are not aligned with the actual needs and challenges of the business. Relevant market and business context needs to be included.

Contd….

The bank took the help of the vendor Fiddler to understand the drivers of the model outputs behind business decisions and to underline factors affecting decision-making . The bank saved money from improved modelling and the resultant decision making. The bank could also understand impact of ineffective model very early and thus preserve valuable resources.

Source : https://www.fiddler.ai/analytics company case study, Accessed on 02/10/2022

Predictive analytics is a technique for predicting the future scenarios for the organization. It gets data from the data warehouse and applies mathematical algorithms to predict the future trends of a business. It uses the techniques such as regression, logistic regression, time-series analysis, duration analysis, situational analysis, multivariate analysis, classification, association, and machine learning techniques such as neural networks and radial basis functions. Let us do an activity (Activity 17.4) on prescriptive analytics of a nationalized bank.

Activity 17.4

Prescriptive Analytics

A nationalized bank in India decided to use analytics for finding customer loan requirements. The board of the organization decided that with analytics capabilities they would like to find out the type of customers and the type of various loans, customers are looking for. Can an alternative loan scheme be suggested to a specific customer? What is the best alternative among the available loans? What can be suggested to the prospective customer? The IT team decided to use big data analytics for this purpose. Suggest the team the type of analytics algorithms which suit their requirement.

Answer:

17.16 Moving beyond the Tools to Analytics Applications

Organizations are using Data Visualization as a way to take immediate action.

The companies are collecting billion rows of data a day and generating tableau for analytics and reporting. Everyday product managers analyze hundreds of millions of rows to understand the user dynamics and problems. Data visualization is making big data analytics iterative. It is also reducing the cycle

time of big data analytics so that immediate action can be taken. Big data visualization is still in the early stages and commercial vendors are using open-source projects who are leading the charge. The following are some open-source projects:

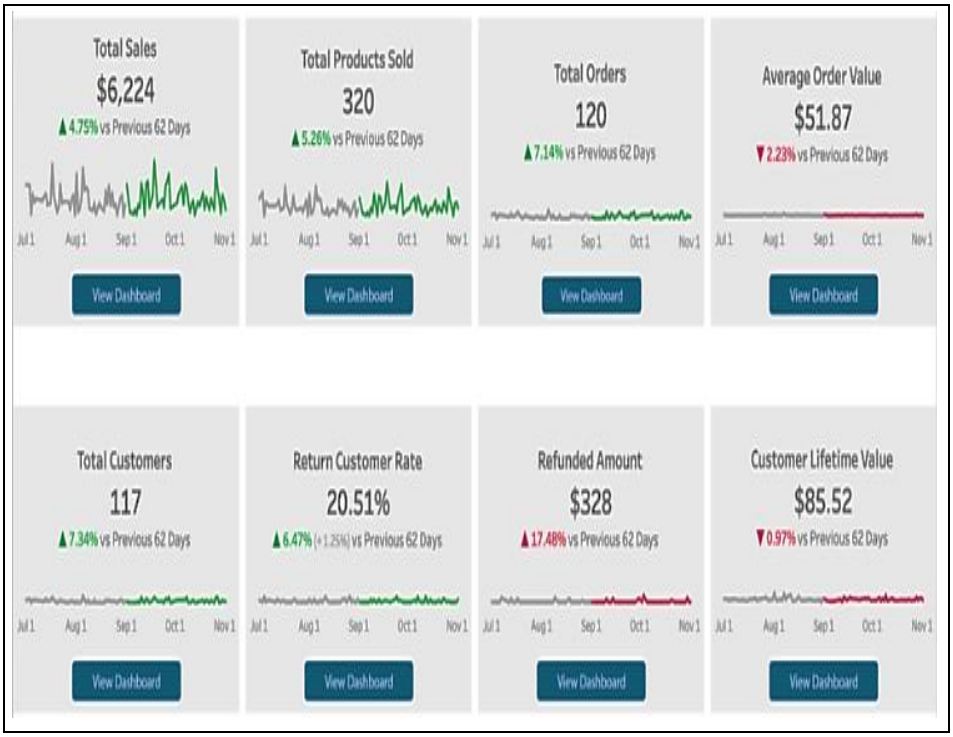
- Qlikview – www.qlikview.com
- Tableau – www.tableausoftware.com
- Micro strategy – www.microstrategy.com
- SAS – www.sas.com
- Cubism (a plug-in for D3 for visualizing time series) – http://cubsim.com
- Arbor JS, a java-based graph library, http://arborjs.org
- Java Script Info Vis Toolkit, http://thejit.org
- Many Eyes, data visualization tools from IBM Research.

Figure 17.5 shows an example social media dashboard and Figure 17.6 shows an example e-commerce dashboard.

Figure 17.5: Social Media Dashboard

Source: Calzon, B. (2022, September 2). See Social Media Dashboards & Tools For Efficient Analytics. BI Blog | Data Visualization & Analytics Blog | Datapine. https://www.datapine.com/blog/social-media-dashboard-template/

Figure 17.6: Ecommerce Dashboard



Source: PenPath. (2021, March 2). Best Ecommerce Dashboards for Data-Driven Growth. https://penpath.com/resources/best-ecommerce-dashboards/

Example: TCS, in Partnership with Qlik, Built A Holistic, Self-Service Knowledge and Insights Discovery Platform for its Employees

TCS has increased the availability of essential information for business operations such as Business Development and Sales, Project Delivery & Managed Services, Talent Management & Resources, Customer Service, Marketing, and Finance & Finance Operations by using over 700 Qlik Sense dashboards. The technology analyses consumer data and provides staff with highly relevant information, such as client advocacy and feedback, in real-time via mobile devices. TCS may use this data to create a 360-degree perspective of the consumer and their experience, make quick choices anywhere, and provide customised services to its customers.

Source: https://datastorageasean.com/news-press-releases/tata-consultancy-services-drives-data-democratisation-qlik may 20,2022 Accessed on 02/10/2022

Activity 17.5

HR Analytics

The HR manager would like to know the impact their organizational culture has on the employee productivity in the organization. For this purpose, he wants to have data-based evidence. He would like to use the organizational

data warehouse maintained by the IT department for this purpose. He seeks the help of IT project manager in finding the impact of organizational culture on employee productivity. The IT project manager tells the HR manager that some tools can be used to do this work. What are those tools? What statistical technique needs to be applied in this case? **Answer:**

17.17 Introduction to Google Big Query, Google Dataflow and Apache Spark

There are many cloud-based big data managing services provided by different companies. Some of the popular ones are Big Query and Dataflow by Google Cloud platform, Spark technology by Apache, etc.

17.17.1 Google Big Query

Big Query is a subscription-based data analytics service by Google Cloud platform to manage big data. The main advantage of using this service is that there is no need to manage IT infrastructure or hire a database administrator. Minimum knowledge of SQL is essential to query the data from the cloud. Organizations with high volumes of real-time data on which analysis is required can make use of this service. Query operations on massive datasets need specialized expensive hardware and is time-consuming; such requirements can be handled using the processing power of Google's infrastructure. There are many Fortune 500 companies and even startups that use these services. Big Query has the following components:

- *Projects:* These are top-level containers in Google Cloud platform used to store Big Query data and information regarding billing and authorized users. Every project has a unique ID.
- *Tables:* The tables contain data in Big Query, schema with field names, data types details.
- *Datasets:* The datasets allow to organize and control access to your tables; tables are entities within the datasets.
- *Jobs:* Jobs is a Set of action tasks like load, export data, query data, or data which is executed by Big Query. Jobs execute independently and they consume time.

17.17.2 Google Dataflow

Google Cloud Dataflow is a tool to perform data-processing tasks on data, irrespective of its size and type.

Cloud Dataflow consists of two major components:

- It consists of a few SDKs which allow defining data processing jobs. Dataflow SDKs are based on a unique programming model to handle large-scale cloud data processing. Data processing jobs are defined by writing programs with the help of Dataflow SDKs.
- The Dataflow service integrates a set of Google Cloud Platform technologies, like Google Compute Engine, Google Cloud Storage, and Big Query. These are used to execute data processing jobs on Google Cloud Platform resources.

17.17.3 Apache Spark

Apache Spark is a fast open-source cluster computing framework used for big data processing, with a built-in library to support streaming, SQL, machine learning, and graph processing activities. This framework is maintained at AMP Lab at UC Berkeley. Compared to Hadoop, it has dual stage MapReduce method. Apache Spark provides much faster performance for some special applications. The major advantage is its ease of use. Due to the availability of high-level operators, parallel apps can also be built. Its high-performing tools like Spark SQL, MLlib for machine learning, GraphX, and Spark Streaming allow the user to perform streaming and complex analytics activities.

Example: Schneider Electric built an Internal Solution for Data Automation and Workflow Automation using Apache Airflow and Apache Spark

Schneider Electric has built an internal solution for data automation and workflow automation using Apache Airflow and Apache Spark. The solution facilitates automation of data science development workflows starting from coding, data processing to machine learning training, and deployment platforms.

Source: https://analyticsindiamag.com/the-companies-that-won-data-science-excellence-awards-2022/ September 29, 2022, Accessed on 02/10/2022

Check Your Progress - 4

11. Which of the following organizations provide business intelligence tools?
    a. Business Objects
    b. Microsoft
    c. PeopleSoft
    d. Business objects, Microsoft and PeopleSoft
    e. Wipro

12. Which of the following is not a big data characteristic?
    a. Size
    b. Speed
    c. Data type
    d. Data Variety
    e. Cost
13. Hadoop features include which of the following?
    a. HDFS (Hadoop Distributed File System)
    b. MapReduce
    c. HDFS and MapReduce
    d. High-Cost Solution
    e. Storage
14. Big data analysis process does not include which of the following?
    a. Data Acquisition
    b. Data Integration
    c. Data Analysis
    d. Inconsistency
    e. Data Restructuring

17.18 Summary

- The challenges facing business intelligence include the volume of data (size), security, data retention, performance targets, and benchmarking. However, business intelligence systems are useful for strategic, tactical and operational planning and decision-making.
- Data warehouse is the major component of business intelligence. It helps in the propagation of data in the organization. It extracts, cleanses, integrates, transforms, and stores the data and further transmits it for query processing and analysis.
- The essential components of business intelligence systems include data warehouse, data marts, corporate performance management, ETL tools, OLAP, analytical tools, data visualization, data mining, geographic information system, and a well-defined workflow.
- Business intelligence tools include AQL (Associate Query Logic), decision support systems (DSS), executive information systems (EIS), management information systems (MIS), query and reporting tools, OLAP tools, data mining tools, and ETL tools.
- There are many cloud-based big data managing services provided by different companies; some of the popular ones are Big Query and Dataflow by Google cloud platform, Spark technology by Apache.

17.19 Glossary

Customer Life Time Value: It is a prediction of the net profit attributed to the entire future relationship with a customer**.**

Data Cleansing: It is the process of removing errors in the data. Incomplete, erroneous and inconsistent data is removed from the data. This process is known as data cleansing. It is required after data acquisition.

Expert Systems: These are the systems developed using artificial intelligence techniques. They can be developed using rule-based programming language. For example, Prolog (Programming in Logic) can be used to develop expert systems. Expert systems can be used in medical diagnosis as well.

Legacy Systems: These are the systems developed using first-generation programming languages. The extension and reusability of these systems were very complex. They are not extendable like object-oriented systems. They were developed using programming languages such as BASIC and FORTRAN.

Metadata: It is the data about the data. The date and time of data creation, who created the data, size of data, the date and time last modified, who has recently accessed the data, etc., are maintained in the metadata. There are metadata repositories in business intelligence systems.

Pivot Analysis: Pivot analysis is done to find out the projected support and resistance levels of data. It is used to generate ad-hoc reports useful for managerial decision- making.

Taxonomy: It is a technique of analyzing semi-structured or unstructured data. Usually, classification is applied to semi-structured data analysis.

Transaction Processing Systems: These systems were developed using business-oriented programming languages such as COBOL. They used to have a master file and a transaction file. They also support batch processing instead of online processing. These were used in banks in the early days of computing for transaction processing. Now, these transaction processing systems are replaced by online distributed systems and cloud computing in multi-national banks.

17.20 Self-Assessment Test

1. Distinguish between data, information and knowledge. How are business intelligence systems helpful in knowledge management in the organization?
2. What are the features of a data warehouse? Explain.
3. Explain the business intelligence architecture. What are its main components?
4. Briefly describe the data mining techniques. Which techniques can be used to find associations between two events?
5. Write a short note on Google Big Query.
6. What is big data? What are its characteristics? What are the important features of Hadoop?

17.21 Suggested Readings/Reference Material

1. Maleh, Yassine. Shojafar, Mohammad. Alazab, Mamoun. Baddi, Youssef. Machine Intelligence and Big Data Analytics for Cybersecurity Applications (Studies in Computational Intelligence, 919) 1st ed. 2021 Edition.
2. Ahmed, Syed Thouheed. Basha, Syed Muzamil. Arumugam, Sanjeev Ram. Patil, Kiran Kumari. Big Data Analytics and Cloud Computing: A Beginner's Guide, 2021.
3. Saleem, Tausifa Jan. Chishti, Mohammad Ahsan. Big Data Analytics for Internet of Things 1st Edition, April 2021.
4. Jones, Herbert. Data Science: The Ultimate Guide to Data Analytics, Data Mining, Data Warehousing, Data Visualization, Regression Analysis, Database Querying, Big Data for Business and Machine Learning for Beginners Hardcover – 10 January 2020.
5. Maheshwari, Anil. Data Analytics Made Accessible: 2023 edition Kindle Edition
6. Mayer-Schönberger, Viktor. Cukier, Kenneth. Big Data: A Revolution That Will Transform How We Live, Work, and Think Paperback – October 26, 2021.

17.22 Answers to Check Your Progress Questions

1. (c) Organizational performance and decision-making.

Business intelligence is required for both organizational performance and decision-making.

2. (c) Data warehouse and data marts.

Business intelligence systems make use of data warehouse and data marts.

3. (c) Processed data

Information is processed data.

4. (d) Decision Support System

DSS stands for Decision Support System.

5. (a) Understanding

Knowledge depends on understanding.

6. (c) Structured data or unstructured data

Data can be structured data or unstructured data.

7. (d) Restructuring Data

Data warehouse functionality includes cleansing data, storing data and transforming data.

8. (d) Data warehouses, data marts, and ETL tools

Business intelligence architecture consists of data warehouses, data marts and ETL tools.

9. (c) Scalable and Secure

Business intelligence architecture should be scalable and secure.

10. (a) Extract, Transform and Load

ETL stands for Extract, Transform and Load.

11. (c) Online Analytical Processing

OLAP stands for Online Analytical Processing.

12. (d) Data warehouse, specific to business function, specific to department

Data mart is a sub-set of data warehouse, specific to the business function, specific to the department.

13. (b) Data Marts

Cubes are derived from data marts.

14. (d) Predictive analysis, statistical analysis and ad-hoc analysis

Business intelligence methodologies include predictive analysis, statistical analysis and ad-hoc analysis.

15. (e) Factoring

Classification, multidimensional analysis and clustering are data mining techniques.

16. (d) Business objects, Microsoft and PeopleSoft

Microsoft, PeopleSoft and Business Objects provide business intelligence tools.

17. (e) Cost

Big data characteristics include size, speed and data type.

18. (c) HDFS and MapReduce

Hadoop features include Hadoop Distributed File System (HDFS) and MapReduce.

19. (d) Inconsistency

Big data analysis procedure includes data acquisition, data integration, data analysis, etc.

Unit 18

Data Privacy and Ethics

Structure

18.1 Introduction

18.2 Objective

18.3 Data Privacy

18.4 The Privacy Landscape

18.5 Database Marketing and CRM

18.6 Rights Come with Responsibility

18.7 Global Privacy Principles

18.8 Personal and Protected Information

18.9 Conscious Responsibility

18.10 Is Privacy the Right Area to Focus?

18.11 Anonymizing Data

18.12 Privacy and Counter-Terrorism Measures

18.13 Data Privacy and Future Course of Action

18.14 Summary

18.15 Glossary

18.16 Self-Assessment Test

18.17 Suggested Readings/Reference Material

18.18 Answers to Check Your Progress Questions

"As our own species is in the process of proving, one cannot have superior science and inferior morals. The combination is unstable and self-destroying."

- Arthur C. Clarke (1917 -), English
Physicist and Science Fiction Author

18.1 Introduction

Data Analytics as a technology is processing huge volumes of data to provide very useful insights for decision-making. But ensuring the privacy of the data is a responsibility.

In the previous unit you are exposed to Need for business intelligence, data warehousing, business intelligence components, business intelligence architecture, business intelligence methodologies, data mining techniques, market intelligence and decision making. We have also learnt about business

intelligence tools like Hadoop, Google Big Query, Apache Spark and Google Dataflow.

As data explosion continues due to increased usage of digital devices and growing online user base, huge volumes of data are produced each day. Companies are collecting data on consumers to understand them and serve them better. Database marketing has paved way to customer relationship management (CRM) system. This includes the personal information of users and sometimes sensitive information too. The important question to understand is who owns the data? Where does it reside, in what format? Under what all contexts could this data be used? In this unit we will know about data privacy and ethics and various rights and responsibilities in connection with data privacy and ethics. Data privacy laws gained prominence in the context of misuse of data & information, leading to harmful effects for the society as a whole. The responsibility to maintain the privacy of data is with the organizations which collect it as those who will be subsequently using it. Rights do come with responsibility. While companies can collect information with the consent of consumers, it is the responsibility of these companies to maintain the privacy of data, besides ensuring that it is not misused. Even for the collection of data, certain rules are to be followed depending on the law of the land. Developed nations have some agencies in place such as Federal trade commission in the US which deals with ensuring data privacy and protecting the consumers' rights on choice and consent.

18.2 Objectives

After going through this unit, you will be able to:

- Discover importance of data privacy
- Compile privacy landscape and the prevalent data privacy laws
- Relate to global privacy principles
- Express personal information and types of protected information
- Describe contextual integrity and conscious responsibility
- Explain anonymization of data

18.3 Data Privacy

With the increase in internet penetration and growing social networks, there is a huge growth in online activity of users. On an average, each person on earth has approximately 2.5 social network connections. As more and more users are using digital devices and are having online presence, it becomes imperative as to how responsible they are in using this online media and what information they are sharing in the public domain, in exchange for accessing the services provided by various social networks, websites, and other online media. It is akin to having a mutual agreement of exchange of value between online users and online service providers. The 'Terms and conditions' of such agreements and interpretation of

them hold importance when it comes to using online services conforming to law while maintaining ethical values. In personal relationships, it is common that the two parties negotiate on the common terms and conditions. However, in case of online/digital services, generally the user either accepts the terms and conditions or simply avoid the service. Usually, the negotiation of terms does not exist in case of online services.

Most of the users think that it is hard for someone to analyze their online presence or activities, but, the advances in computational power and analysis has made it easy to analyze private lives. It has become harder to keep secrets of personal lives. Users frequently update their status, post photographs, comment, tag other people, mention the places of visit etc. This can easily provide some insight into the lives of users for someone who is analyzing.

Example: TikTok was Levied $29 million for Failing to Protect "Children Privacy"

TikTok was levied a penalty of $29 million by U.K.'s Information Commissioner's Office (IOC) for failing to protect "child privacy" for two years. IOC found that the company might have processed the data of children under 13 without the consent of parents. IOC alleged that TikTok did not inform the children and parents about data privacy terms in a transparent manner. This is especially critical in the case of special category data such as sexual orientation, religious beliefs, ethnic and racial origin, political opinions, and genetic and biometric data.

IOC felt that children should get access to the digital world, the companies need to take enough care to protect the data privacy. The data could go into wrong hands and the children are vulnerable.

Source: TikTok faces $29M fine in UK for 'failing to protect children's privacy' | TechCrunch date 26/09/2022, Accessed on 03/10/2022

18.4 The Privacy Landscape

The four important constituents of privacy landscape are mentioned below:

- **Organizations**: It is very important for businesses and organizations to leverage the intelligence information that is personally identifiable, sensitive, and competitive to their advantage. They have to make appropriate investments in information systems, data procurement, and analysis to drive decisions.
- **Misuse of data by criminals**: As personal information of users is available online, impersonation, identity theft, stealing confidential information has surged. Hackers are using sophisticated technology to exploit the data security loopholes to steal data.

- **Online users/customers**: Recently there has been increased awareness and concern of online consumers about the collection of their data and the use of it.
- **Law makers:** Consumers' concerns need to be considered and the access to use personal information has to be restricted.

Example: India is Ready with the New Data Privacy Act to make the Online World More Accountable

Indian telecom minister Vaishnav has indicated that the new "data protection bill" is ready to be introduced in the parliament and that would make the online world more accountable. The Centre withdrew the Personal Data Protection Bill, 2019 in August 2022. The withdrawn data bill had proposed restrictions on the use of personal data without the explicit consent of citizens. It had also proposed to give government agencies exemptions. This move was opposed by Opposition parties. The new proposed legislation will address all concerns as per the Minister. The Government feels every citizen, organizations, Government agencies all have a role to play.

Source: Nirmala Sitharaman - Data privacy bill ready for launch - Telegraph India dated 8/9/2022, Accessed on 03/10/2022

18.5 Database Marketing and CRM

Historically, the collection of information about consumers is done directly from consumers as well as from third-party players. This data collection is already in place for the past few decades. Previously, for products sold, the primary means of data collection was in the form of warranty cards. Providing warranty cards usually comes with having the customer profile collected as a pre-requisite, and this eventually served as a repository/database of customers for companies. More often than not, this process of providing cards and collection of customer information is handled by third party players. They collect information from customers of various companies and sell the bundled information to other companies.

Slowly, the collection of information about customers led to the growth of database marketing. The advent of advanced information systems and analytics techniques has helped marketers understand better; the preferences and choices of consumers, and their channel response. Companies also realized the importance of using big data to design products or services to better suit consumer needs. Until late 1980s, specialized players used to handle collection of customer data, and provided companies with segmented lists. However, this process was tedious and time consuming. By the time companies receive the segmented lists, a whole set of new data gets produced. In 1990s, companies started having their own databases for maintenance of customer data and for segmentation. Advances

in analytics field has reduced the processing time from several months to a few days, and currently the trend is moving towards near real-time processing.

18.5.1 Consumer Data and Customization of Products &Services

Consumers are willing to share personal information to companies in exchange for enhanced value. Organizations are moving away from segmentation approach to customization and personalization. Consumers also prefer to have companies understand them individually and tailor the services specifically to them. Users will not experience the explicit value added for them if they are simply treated along with all other consumers. Database marketing has paved way for CRM (customer relationship management) systems. If the personal information is used responsibly by users as well as organizations, that will be mutually beneficial for both the parties. Ideally, we, as consumers, would like the organizations to have the entire information about us across all products, all purchases made, by connecting our conversations and relationships with them. At the same time, consumers may not like the idea of some organizations asking for personal information and details even before consumers decide to purchase products and services from them. While lack of personalization is a setback for organizations, focus on privacy policies and protecting the privacy of customer information is vital for the relationship between them and customers.

Example: IRCTC Plans to Monetize Customer Data (CRM Data Base) Opposed Due to Privacy Concerns

IRCTC has floated a tender inviting a consultant to suggest ways to monetize its customer data. As per the company, the objective is how to offer better services to its customers based on the huge amount of data collected over time. The data covers both passenger and freight operations. But advocacy groups are resisting this saying the company would violate privacy issues to generate additional revenues. The company may withdraw the tender until the new privacy laws are in place. The data to be analysed will include name, age, mobile number ,gender , address, email id, class of travel , payment mode etc. The company has over 10 crore customers in its CRM database and around 7.5 crores are active.

Source: IRCTC floats tender to monetise customer data, may withdraw over privacy concerns | India | Onmanorama date 22/08/2022, Accessed on 03/10/2022

18.6 Rights Come with Responsibility

In this growing online & digital world, consumer privacy is a serious matter of concern. Consumers' rights and choices are to be protected. There is an absolute necessity to protect consumers from abuse of companies' access to their private data. Let us take a closer look at the laws and regulations around the consumer privacy in the US, one of the most developed nations in the world, where internet penetration and digital usage is quite high.

18.6.1 Do-Not-Track Bills

This term refers to the collective series of policy proposals that aim at protecting consumers' right to choose whether or not their online activity can be tracked by websites. Private and personal information has become an important tool for companies and organizations to design effective marketing strategies and implementing them. However, there is a growing concern on consumer privacy and many US legislators are trying to enact laws that protect privacy of internet users. Even before internet was heard of, in 1986, ECPA (Electronic Communication Privacy Act) has set limitations to access private information through law enforcement. The Children's Online Privacy Protection Act (COPPA) introduced in the year 2000 prevents tracking of online activity of all children under thirteen years of age and personal information of children cannot be gathered without the consent from parents.

In 2010, the US FTC (Federal Trade Commission) has published a report on consumers' right to prevent websites from collecting private information and to track their online behavior. In relation to this, there were several bills passed such as, Do-not-track-me-online Act of 2011, and Do-not-track-kids Act of 2011. HIPAA (Health Insurance Portability and Accountability Act) was introduced in 1996 for protection of confidential medical information. This act requires organizations, healthcare providers, as well as their business associates to develop and to follow, the processes and procedures that ensure security and confidentiality of PHI (Protected Health Information), while transferring, receiving, handling or sharing it. Also, only the minimum health information required for business has to be used. The GLBA (Gram Leach Bliley Act) also known as the Financial Services Modernization Act; introduced in the year 1999 requires companies, banks, and financial institutions that offer consumers with financial products and services including loans, insurance services, investment advice etc. to clearly explain their information sharing practices to consumers. They also need to safeguard the sensitive data and information.

18.6.2. Role of Federal Trade Commission

FTC (Federal Trade Commission) was created way back in 1914 when Federal Trade Commission Law was signed by the then US president. It came into effect in 1915 with an aim to protect consumers and promote competition among companies and organizations. FTC develops tools, processes, and policies through workshops & conferences and partners with law enforcement agencies to advance consumer protection and competition missions. FTC has a set of guidelines on online privacy that it enforces to advance online privacy. While online privacy is a matter of exchange between organizations and consumers, it is mainly self-regulated, although some state laws exist to protect it. For example, California office of privacy protection requires companies, persons, and agencies that operate business in California and have access to digital 'personal information' to report any breach of security. Apart from HIPAA and GLB

discussed in earlier sections, legislative action can also be taken in cases where FTC requires a company to honor its stated privacy principles and policy.

Example: Tilting Point media LLC (Owner and Operator of Spongebob App) Found Violating Children's Online Privacy Protection Act ("COPPA")

Tilting Point Media LLC is a mobile-game publisher. The Company offers brand marketing, copywriting, game event management, lifecycle marketing, gaming funding and other related services using artificial intelligence (AI) driven marketing technology. Tilting Point Media serves customers worldwide.

The Children's Advertising Review Unit ("CARU") found that Tilting Point Media, LLC ("Tilting Point"), owner and operator of the SpongeBob: Krusty Cook-Off app (the "App"), violated the Children's Online Privacy Protection Act ("COPPA") and CARU's Self-Regulatory Guidelines for Advertising and for Children's Online Privacy Protection ("CARU's Guidelines"). CARU has recommended many corrective actions to be taken up by the company towards data privacy.

Tilting Point is expected not to collect any information from children of age below 13. Even if it wants to collect specific information, parent consent is to be obtained. It was observed that the age was not checked in the app.

Source: SpongeBob App Found to Violate Child Online Privacy Laws (natlawreview.com) 26/09/2022, Accessed on 03/10/2022

18.7 Global Privacy Principles

In an attempt to be more transparent and to build trust & relationship with consumers, most companies come up with their own set of rules guiding their privacy policy. The core of these rules and principles that companies develop are mainly based on 'EU-US Safe Harbor Principles'.

The seven global privacy principles are outlined below.

1. **Notice/Transparency** – Individuals must be informed by companies on the process and purpose of collecting information from them and the way it is being used. Also, information must be provided on how companies can be contacted for any queries or complaints.
2. **Choice/Consent** – Individuals should have the choice to opt out from their data being collected and used by company/third party. They should have an option to choose how the personal information they provide will be used.
3. **Transfer of Data** – This rule pertains to sharing of data with other organizations or vendors or '-third party service providers. Transfer or sharing of data by companies to others can be done only if the other parties are conforming to data protection principles.

4. **Security of Information** – Companies should take responsible measures to curb any data theft or misuse of collected information including alteration, destruction, manipulation etc. Reasonable and substantial measures are to be taken in this direction.
5. **Data Integrity** – Assuring that the information and data collected is reliable, and steps or measures are taken to ensure that the information collected is relevant, current accurate and complete.
6. **Access** – Individuals should be provided access to their personal information and should be able to correct or modify it as needed.
7. **Accountability/Enforcement** – Organization or company must be responsible and accountable for following these principles and should have mechanisms to ensure compliance to these rules.

Example: HomeLane.com has Zero Tolerance Transparent "Data Privacy" Policy

HomeLane is a technology enabled home interior company in India. It provides customized home interior solutions to its customers with a transparent pricing policy. The leadership team also believes that Data Privacy is paramount to the growth and sustainability of its business. Also, it considers Data Privacy is a KRA (Key Results Area) for the top leadership and at the organizational level. It has a transparent privacy policy in place, and it is implemented with regular audits and management reviews.

The company believes with power comes responsibility and does not want short time gains through data monetization.

Source: How HomeLane utilises AI (analyticsindiamag.com) 06/04/2022, Accessed on 03/10/2022

Activity 18.1

Take a company website of your choice with free registration and register as a user for it. Check out the 'Terms and conditions' section. What points and details are covered in this section? Is there any mention of data privacy and the measures they take to ensure data privacy?

Answer:

Check Your Progress - 1

1. Which of the following holds importance for data privacy while maintaining ethical values?
   a. Digital devices
   b. Online presence
   c. Responsible use of online media
   d. Information shared in the public domain
   e. Terms and conditions
2. Which of the following is not an important constituent of privacy landscape?
   a. Organizations/companies
   b. Criminals misusing data
   c. Law makers
   d. Online consumers/users
   e. Profits generated by company
3. Which of the following is not a global privacy principle defined in the 'seven global privacy principles'?
   a. Consent
   b. Data Integrity
   c. Accountability
   d. Mutual implicit trust
   e. Access of information/data
4. Who handles the process of collection of information about consumers?
   a. Consumers
   b. Third-party players
   c. Marketers
   d. Individuals
   e. Users

18.8 Personal and Protected Information

The definition of personal information has changed over time. Decades ago it was just the first name, last name, address and phone number that primarily constituted the definition. Over time, this has changed and included email address, facsimile number, and also unique identification numbers issued by governments. Even the static IP address, Cookie Ids and financial account numbers can be unique and can be mapped to a single individual. Apart from personal information, there is also something called as sensitive information. The collection and use of it requires more attention and care. For example, sexual

orientation information of an individual is not just personal information but it is also sensitive. Besides, this can lead to harmful results and embarrassment to individuals, if misused. Although many companies have data policies in place to handle sensitive data, much of this focuses on sensitive data of company itself and not on the data collected by them. Hence, it is vital to have data management policies that include differentiating various types of data (typically called data classification)

18.8.1 Protected Information

This can be classified into three groups: a) Personally identifiable information, b) Sensitive information, and c) Other information (non- identifiable information about individuals used when in combination with PI – personal information).

Let us look at the details of each of these types to gain more insights on this classification:

- a) Personally identifiable information
  - First name, last name
  - Email address, mobile number
  - Social security number (SSN is specific to US, could be Aadhar number in case of India), Driving license number
  - Account information of individuals in banks and financial institutions, Credit card and Debit card information
- b) Sensitive information – Disclosing this information could demean or embarrass an individual
  - Ethnicity, Race, Gender, Age, Religious beliefs or opinions
  - Marital status, sexual orientation
  - Membership with trade unions
  - Criminal history/record
  - Disease information/other health information
- c) Other information
  - Static IP addresses
  - Cookie IDs
  - Preferences and choices

Activity 18.2
Check the associated information/data that you frequently carry along with you (ID card, Debit card, Credit card, Driving license, etc.) Classify each of these pieces of information into personal or protected information types.

18.8.2 Contextual Integrity

The context in which particular information is collected holds importance as the information may be relevant to a particular situation or stated purpose, but is used for some other purpose and in some other situation. CI (contextual integrity) refers to the alternative benchmark developed to evaluate breach of privacy. Society has different settings and contexts which have to be acknowledged and considered while evaluating potential privacy breach. For example, the setting of a hospital is pretty much different from that of an academic institution. Hence, contextual integrity enables individuals to define or describe their expectations of privacy by associating various types of behaviors with contexts. Information and data is associated with contexts, so, the data is considered sensitive only with respect to the associated contexts.

Example: Apple has Designed a Privacy Preserving Approach to Protect Personal Data and at the Same Time Getting Insights about the Customer Preferences

Apple has developed a technique known as "local differential privacy" whereby Apple obtains customer insights while preserving the individual privacy. Through this, Apple learns about the user community without learning about individuals. Differential privacy transforms the information before it is sent from the user's device. This ensures Apple cannot reproduce the original data.

Source: Differential Privacy Overview (apple.com), Accessed on 03/10/2022

18.9 Conscious Responsibility

Although there are some major incidents reported on privacy failures, one has to acknowledge significant efforts from companies and organizations to conform to the data protection and privacy policies. Many companies are conscientious of their approach in designing their privacy policies, issues related to privacy and cases where there is potential violation of individual rights. They are aware of the consequences of data privacy violation and are taking steps to prevent it. Perhaps, the best way to manage sensitivity issues is to enable individuals to choose what is suitable to them, in addition to providing them with adequate and active feedback loops, where the dialogue exchanged and feedback becomes crucial for operational changes.

Example of Data Privacy Violation

In February 2012, Arun Thampi, a developer from Singapore has found that an iPhone app was secretly downloading the address book of users without their consent. 'Path' is an iPhone app that is used for sharing as well as journaling the

paths of one's own life. Once the app is opened and registered, it automatically uploads the contact data or address book of users, to find friends that they might want to connect with.

This is a clear example of violation of data privacy and this led to the public apology from the CEO of 'Path', Dave Morin. Immediate action was taken to delete all the collected information from their servers. Software of 'Path' was fixed so as to give users/individuals the control of such data. In the later versions of this app, users are prompted either to opt in or opt out from sharing the contacts in their phone with the servers of 'Path'.

Exhibit 18.1 discusses data breach.

Exhibit 18.1: One of Biggest Data Breaches of the 21st Century

China's Sina Weibo (SW) an alternative to Twitter, having over 500 million subscribers, reported in March 2020 that the user details such as the real names, site usernames, gender, location, and phone numbers of 172 million users were put for sale on dark web markets.

Weibo acknowledged the sale but claimed the data was obtained by mapping contacts with phone numbers drawn from address book API. Weibo assured it users not worry as none of the passwords are stored in plain text format. The social media giant notified authorities on the incident and China′s Cyber Security Administration of the Ministry of Industry and Information Technology is investigating the data privacy violation issue

Source: Swinhoe Dan (2021) The 15 biggest data breaches of the 21st century. CSO India (January, 2021)

https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html

Example: Clearview AI Fined $8 Million for Violating the Data Privacy of UK Residents

Clearview AI is an American facial recognition company, providing software to companies, law enforcement, universities, and individuals. The UK Information Commissioner's Office (ICO) has fined Clearview AI Inc. $8 million for violating the data privacy of UK citizens.

ICO ordered the company to stop obtaining and using the personal data of UK residents available on the internet and to delete the data of UK residents from its servers.

The company collected highly sensitive biometric information without the consent of individuals.

Source: Clearview AI fined over $8 million for data privacy violation | Security Magazine 31/05/2022, Accessed on 03/10/2022

18.10 Is Privacy the Right Area to Focus?

There has been good emphasis on data privacy and protection of information from the perspective of law and also as a policy. However, sometimes it is not about data privacy but it is a matter of trust and relationship that is maintained with customers, so, it boils down to ethics followed by companies and organizations, rather than simply having privacy policies in place.

There are always some exceptions to the general rules of privacy that information should not be shared. For instance, in emergency situations, disclosing information may be quite helpful. In order to accommodate such scenarios, there are exceptions to the requirements of giving notice and receiving consent from users. In some cases, it is necessary to collect some information, but, this needs to be done ensuring that this information is not identifiable at individual level.

Example: Headspace Mandated Its Research Collaborators to Adopt Ethical Practices while Dealing with Customer Data

Headspace, a subsidiary of Headspace Health, is an English American online company, specializing in meditation. The company works with research organizations to analyse the vast data it has collected from its clients to get insights into effectiveness and impact of meditation on various situations . The company believes trust and ethics are of paramount importance. It mandated its collaborators to strictly ensure the privacy of data by obtaining the consent of the users after explaining the purpose for which the data is used.

Source: Headspace calls for third-party research collaborators (fiercehealthcare.com) date 29/09/2022, Accessed on 03/10/2022

18.11 Anonymizing Data

Data anonymization is a technique of information-sanitization where the process of either encrypting or removing the PII (personally identifiable information) from original data sets is followed, so as to mask the identity of individuals in original information. This is generally done to conform to data privacy guidelines.

By masking the identity of individuals behind the original information, data collected for a particular purpose can be used for other purposes (termed as 'secondary use'), by stripping the data of the classification of PII (personally identifiable information). Also, in some cases where the inferences have to be made at general level and not at individual level, data can be anonymized and analyzed. For example, if the objective of analysis is to analyze the trends of a disease at regional level, original identity of individuals is not a factor of importance, if the data on region of user is collected. However, the flipside of anonymizing data is that research sometimes requires real data about real people. Perhaps, that could turn out to be a life-saving breakthrough research, which cannot be performed on masked data.

Example: New Microsoft Analytical Tools help Understand Trends Without Violating Data Privacy

Traditionally companies and users thought "anonymizing data" will lead to perfect data privacy. But it has been found that it is not true. Anonymizing is time-consuming and not completely reliable. Microsoft identified a better way by creating synthetic data. Synthetic data has all the properties as the real data except that the data cannot be leaked. Synthetic data to be useful has to match the distribution of the original data set, down to the combinations of individual characteristics.

Source: New Microsoft analytics tools help identify and understand trends without compromising privacy | TechRepublic 23/09/2021, Accessed on 03/10/2022

18.12 Privacy and Counter-Terrorism Measures

While data privacy advocates more and more security, it becomes a matter of concern when scenarios like counter-terrorism demand otherwise. While data privacy, anonymization, and encryption help companies prevent hefty fines, encryption could pose a problem to governments and security agencies. Let us take a look at the case below, where data privacy battles with counter-terrorism measures.

Case Study 18.1: Data Privacy vs. Counter-Terrorism Measures

While more encryption brings in more security and protects data privacy, too much of encryption may not be good in all cases. Privacy advocates demand more security and encryption on one hand, while the US government is worried that terrorists take refuge under the platform of encrypted communicating system.

Recently, an attack in California has led investigators (FBI) to look into the possible information that could be retrieved from the iPhone of the deceased attacker. Ten unsuccessful attempts to provide correct password would lock investigators of access to information in phone. Apple's messaging system uses complete (end to end) encryption which means the text messages are encrypted on sender's device that could be decrypted only on receiver's device.

A court order has sought Apple to help FBI and government by building a backdoor to circumvent the encryption and get access to information in the attacker's phone. Apple is however firm in following its stated data privacy policies and said it will not create 'backdoor' software to unlock its phones.

Apple CEO, Tim Cook mentions that backdoors not just allow 'good guys' like law enforcement to use it but also helps 'bad guys' to access private customer information that may include critical information such as bank and healthcare records.

Source: ICFAI Research Center

18.13 Data Privacy and Future Course of Action

As big data emerges, collection of data from users has become increasingly relevant to improve business. Equal importance has to be accorded to data privacy and protection policies. Rather than looking at it purely as a compliance matter, data privacy should be taken as an opportunity to build trust and relationship with consumers.

There may not be a fool proof approach or strategy for maintaining data privacy. Companies have to regularly revisit the policies in place to accommodate to the new set of challenges, as new contexts and issues will continue to emerge with evolving technology.

Case Study: Privacy Rights Violation by Google

Google, the world's largest search engine was found to have violated privacy laws. It was discovered that Google was implanting some cookies on Apple's safari browser (a prominent browser of choice on iPads and iPhones) to bypass privacy settings. This helped the search giant to access the online buying behavior and online browsing habits tracked using third party cookies.

Google had a feature that puts +1 button in advertisements placed on the web using its DoubleClick ad technology. If users like the advertisement, they would click "+1" and this serves as an approval to their social networking profile of Google. However, safari by default blocks most of the tracking. This disabled Google to use their 'cookie' installation technique to check if the safari users were logged into Google.

Google had signed a consent decree with FTC to settle allegations that it has used deceptive techniques to garner information from online users violating its stated privacy policies. The 20 year agreement that Google has signed bars it from misrepresenting the way it collects & handles user information.

In North America, Google had to pay USD 40 million over this incident after FTC has probed, and several other regulators from 38 states fined it. California based consumer watch dog, Santa Monica, has commented that cookie settlement will still allow Google to use the data it collected in deceitful manner. The internet giant has collected data from 190 million consumers whose privacy settings were bypassed.

To circumvent this default blocking, Google has exploited a loophole in the privacy settings of safari browser. Safari usually blocks the tracking, but makes exception for websites that interact with customer in some way. For example, where customer is filling and submitting a form on a particular website. Google has added some code to the ads placed, that make safari think that the user is submitting a form to website and hence allows Google cookie to get installed.

Source: ICFAI Research Center

Exhibit 18.2 narrates a penalty payment for privacy concern.

Exhibit 18.2: Facebook to Pay $5bn to Settle Privacy Concerns – The Biggest Fine Imposed Ever

As per a BBC report in 2019, The US Federal Trade Commission (FTC) asks Facebook to establish an independent privacy committee with Mark Zuckerberg (Facebook CEO) not have any control over it.

The Probe

The above authority started investigating Facebook in March 2018 following the finding that personal data was illegally acquired through an online personality quiz and was sold to Cambridge Analytica (CA) - a data analytics firm. Cambridge Analytica also offering a political consultancy managed to get hold of 87 million Facebook user profile details – a clear violation of data privacy concerns.

The Violation

1) Facebook quiz invited users to find out their personality.
2) The Facebook app collected 27000 participants data along with their friends data, with any consent.
3) As near 300000 people installed the app but it gathered 87 million users data.
4) Facebook claimed the partially sold to Cambridge Analytica (CA)
5) CA denies the use of Facebook user data during US presidential election.
6) Facebook notified it users informing weather their data breached.

Source: BBC News (2019). Facebook to pay record $5bn to settle privacy concerns. https://www.bbc.com/news/business-49099364

Activity 18.3

List out the laws related to data privacy in India, if any. Describe the data privacy landscape in India and compare it with global guidelines.

Check Your Progress - 2

5. Which of the following statements is not true?
    a. Data anonymization is impossible to achieve
    b. FTC deals with data privacy compliance
    c. Some websites ask consumers for consent to collect information
    d. Email address comes under PII
    e. Data privacy standards differ across countries
6. SSN (Social Security Number in US) falls under which of the following class of protected information?
    a. General information for record purpose
    b. Authentic information
    c. Personally identifiable information
    d. Sensitive information
    e. Unidentifiable information
7. Race and Ethnicity fall under which category of information?
    a. Protected information
    b. Public information
    c. Private but unprotected information
    d. Cannot be classified
    e. None of the above
8. Which of the following statements is true?
    a. Law cannot enforce data privacy
    b. Data anonymization cannot be used for medical information
    c. Anonymized data should not be traced back to individuals by anyone
    d. Encryption methods are illegal as they can help illegal activities as well as terrorists
    e. Electronic health records fall under the purview of data privacy
9. Contextual Integrity
    a. Has nothing to do with data privacy
    b. Deals with data privacy and contexts associated with it
    c. Integrity of organizations in financial reporting
    d. Depends on the type pf data
    e. Refers to integrity of users on social networks

10. Conscious Responsibility
    a. Refers to responsibility of consumers while purchasing products online
    b. Responsibility of users to report issues in websites
    c. Responsibility of organizations to drive data privacy
    d. Is irrelevant to data privacy scenario
    e. Responsibility of governments to curb data theft

18.14 Summary

- As data collection, retention and analysis of data has gained prominence in recent years, more and more consumer data resides in the servers of various organizations and companies. It becomes an absolute necessity to have data privacy laws in place as consumers are sharing 'personal' and 'protected information' with various websites.
- Companies are collecting consumer information, analyzing their online habits, using click stream data, and trying to understand their needs, preferences and choices in an attempt to serve them better.
- There is a clear move from the historical segmented approach to target consumers to personalized approach. However, with more and more consumer data residing online in some or other servers, data privacy becomes crucial.
- Data privacy laws vary across various countries. Even within a country, different states may have different laws to enforce data privacy.
- In Unites States, there have been a series of legislations and laws to enforce data privacy. Do-Not-Track Bills are the product of efforts made by legislators to preserve consumers' rights pertaining to choice and consent of whether the customers' online movement can be tracked by the websites.
- There are few instances where data privacy is being violated including some of the most popular companies. However, it has to be acknowledged that every company has the conscious responsibility to avoid data privacy violation, and efforts are being made in this direction.
- Anonymization of data and encryption are powerful techniques to preserve data privacy and maintain confidentiality of people behind the data.

18.15 Glossary

Anonymization: Masking of data so as to de-identify the original individuals to whom the data pertains to.

Conscious Responsibility: Conscious efforts by organizations and companies to ensure data privacy and protection

Contextual Integrity: This concept mentions that data privacy should be seen in association with the contexts. The context for which the data collected, and the context or purpose in which it is used, holds importance according to this concept.

CRM Systems: Customer Relationship Management systems are information systems and processes, for maintaining and managing customer information, and relationship, that are developed with an aim to move from segmented approach to customized or personalized approach by companies to target individuals effectively.

Data Privacy: User information or consumer information that is collected either directly or indirectly has to be stored and used in accordance with privacy and confidentiality. This is called data privacy.

Do-Not-Track Bill: Collective term that refers to set of legislations and acts passed in an attempt to preserve rights of consumers on providing consent to get tracked or not tracked when online.

Federal Trade Commission: It is an independent agency in the US which deals with ensuring data privacy and having healthy competition

HIPAA: Health Insurance Portability and Accountability Act introduced in 1996 is a United states legislation that deals with data privacy and protection of healthcare related information

Personally Identifiable Information: Information that is unique to an individual.

18.16 Self-Assessment Test

1. What is FTC and its role in ensuring data privacy? Provide a real case or example where FTC pro-actively took up the data privacy issue of a particular company and fought against it.
2. Briefly explain Do-Not-Track Bills in US and how they help in preserving consumer interests.
3. What is contextual integrity? Provide an example for this.
4. Classify with examples various types of protected information.
5. Explain data anonymization, its advantages and disadvantages.
6. Describe the seven global privacy principles.

18.17 Suggested Readings/Reference Material

1. Maleh, Yassine. Shojafar, Mohammad. Alazab, Mamoun. Baddi, Youssef. Machine Intelligence and Big Data Analytics for Cybersecurity Applications (Studies in Computational Intelligence, 919) 1st ed. 2021 Edition.
2. Ahmed, Syed Thouheed. Basha, Syed Muzamil. Arumugam, Sanjeev Ram. Patil, Kiran Kumari. Big Data Analytics and Cloud Computing: A Beginner's Guide, 2021.

3. Saleem, Tausifa Jan. Chishti, Mohammad Ahsan. Big Data Analytics for Internet of Things 1st Edition, April 2021.
4. Jones, Herbert. Data Science: The Ultimate Guide to Data Analytics, Data Mining, Data Warehousing, Data Visualization, Regression Analysis, Database Querying, Big Data for Business and Machine Learning for Beginners Hardcover – 10 January 2020.
5. Maheshwari, Anil. Data Analytics Made Accessible: 2023 edition Kindle Edition
6. Mayer-Schönberger, Viktor. Cukier, Kenneth. Big Data: A Revolution That Will Transform How We Live, Work, and Think Paperback – October 26, 2021.

18.18 Answers to Check Your Progress Questions

1. (e) Terms and conditions

Data privacy is akin to having a mutual agreement of exchange of value between online users and online service providers. The 'terms and conditions' of such agreements and interpretation of them hold importance when it comes to using online services conforming to law while maintaining ethical values.

2. (e) Profits generated by company

Profits generated by company do not figure as an important constituent of privacy landscape

3. (d) Mutual implicit trust

Mutual implicit trust is not a principle defined in global privacy principles

4. (b) Third-party players

More often than not, this process of providing cards and collection of customer information is handled by third party players.

5. (a) Data anonymization is impossible to achieve

Data anonymization is possible to achieve and is a well-known technique

6. (c) Personally identifiable information

SSN is for unique identity, this comes under personally identifiable information

7. (a) Protected information

Race and Ethnicity come under protected information (Sensitive information)

8. (e) Electronic health records fall under the purview of data privacy

Electronic health records maintained by healthcare providers contain important medical information about individuals, and this falls under data privacy scope

9. (b) Deals with data privacy and contexts associated with it

Contextual Integrity - This concept mentions that data privacy should be seen in association with the contexts. The context for which the data collected and the context or purpose in which it is used, holds importance according to this concept.

10. (c) Responsibility of organizations to drive data privacy

Conscious responsibility is the conscious effort made by companies or organizations to follow data privacy rules.

Big Data, Cloud and Analytics

Course Structure

Block 1: Introduction and Applications of Big Data	
Unit 1	What is Big Data?
Unit 2	Why Big Data is Important?
Unit 3	Big Data in Marketing & Advertising
Unit 4	Big Data in Healthcare
Block 2: Cloud Computing and Big Data Technologies	
Unit 5	Big Data and Cloud Technologies
Unit 6	Big Data Technologies and Terminologies
Unit 7	Cloud Computing and Big Data Management for Decision Making
Unit 8	Handling Unstructured Data
Unit 9	Information Management
Block 3: Business Analytics	
Unit 10	Analytics
Unit 11	Business Analytics-I
Unit 12	Business Analytics-II
Block 4: Managing Talent for Big Data Analytics	
Unit 13	Talent Management-I
Unit 14	Talent Management-II
Block 5: Data Privacy and Analytics in Various Business Areas	
Unit 15	HR Analytics in HR Planning
Unit 16	Data Analytics for Top Management Decision Making
Unit 17	Business and Marketing Intelligence Using Analytics
Unit 18	Data Privacy and Ethics